PAUL B. HOEBER
MEDICAL BOOKS
69 E. 59th St., N.Y.

E. H. B.

19 2 183

BOSTON MEDICAL LIBRARY

PURCHASED FROM THE INCOME OF THE

OLIVER F. WADSWORTH FUND

Digitized by the Internet Archive
in 2010 with funding from
Open Knowledge Commons and Harvard Medical School

http://www.archive.org/details/manualofmental02whip

MANUAL OF MENTAL AND PHYSICAL TESTS

Part II: Complex Processes

[SECOND EDITION, REVISED AND ENLARGED]

MANUAL OF MENTAL AND PHYSICAL TESTS

In Two Parts

Part II: Complex Processes

A BOOK OF DIRECTIONS
COMPILED WITH SPECIAL REFERENCE TO THE EXPERIMENTAL STUDY
OF SCHOOL CHILDREN IN THE LABORATORY
OR CLASSROOM

BY

GUY MONTROSE WHIPPLE, PH.D.

ASSOCIATE PROFESSOR OF EDUCATION, UNIVERSITY OF ILLINOIS

AUTHOR OF "A GUIDE TO HIGH SCHOOL OBSERVATION," "QUESTIONS IN GENERAL AND EDUCATIONAL PSYCHOLOGY" "QUESTIONS IN SCHOOL HYGIENE"

BALTIMORE, U. S. A.
WARWICK & YORK, Inc.
1915

BOSTON MEDICAL
DEC 24 1915
LIBRARY
13276 Wood

Copyright, 1915
BY
WARWICK & YORK, INC.

PREFACE TO PART II

If it be not thought bad form to preface a volume with an apology, I should like to ask the indulgence of those of my readers who have been so complimentary as to express to the publishers and to me their irritation at this delay of two years or more in the completion of the revised second edition of the *Manual*. The truth is, that the unexpected exhaustion of the first edition of the book found me quite unprepared to rewrite the text at short notice, and that the whole subject of mental tests had meanwhile so expanded as to present a task of no small magnitude to one who would seek to deal at all adequately with the material that had become available.

In this volume, then, as in Part I, the text has undergone extensive revision and alteration. In a number of instances the addition of new materials, of new methods and of new results has been sufficient to alter the complexion of the tests so decidedly as to amount to entirely new presentations of the topics with which they deal. The Kent-Rosanoff Test and the Analogies Test are introduced as totally new material.

On the other hand, I have been compelled, reluctantly, for reasons set forth in the text, to omit consideration of serial graded tests (Chapter XIII). This omission I hope to repair later on by publishing a supplementary volume dealing with Systems of Tests in general. To incorporate this material in the present volume would increase its size unduly and delay its appearance beyond reasonable limits of time.

In addition to the acknowledgments for assistance made in the preface to Part I, my thanks are extended to Miss Margaret Cobb and Dr. H. O. Rugg for valuable assistance in the reading of proof. Other special acknowledgments I have tried to make in the course of the text.

G. M. W.

University of Illinois, April, 1915.

TABLE OF CONTENTS

CHAPTER VIII. TESTS OF DESCRIPTION AND REPORT

Test 31.—Description of an Object ..... [376] 10

Test 32.—Fidelity of Report: *Aussage Test* ..... [383] 10

CHAPTER IX. TESTS OF ASSOCIATION, LEARNING, AND MEMORY

Test 33.—Uncontrolled Association—Continuous Method ..... [410] 44

Test 33A.—Uncontrolled Association—Discrete Method (Kent-Rosanoff Test) ..... [419] 53

Test 34.—Controlled Association—Logical Relations ..... [437] 71

Test 34A.—Analogies ..... [455] 89

Test 35.—Controlled Association: Computation ..... [460] 94

Test 36.—Mirror-Drawing ..... [485] 119

Test 37.—Substitution ..... [499] 133

Test 38.—Memory for Serial Impressions: 'Rote' Memory ..... [515] 149

Test 39.—Memory for Ideas: 'Logical' Memory ..... [570] 204

CHAPTER X. TESTS OF SUGGESTIBILITY

Test 40.—Suggestion by the Size-Weight Illusion ..... [589] 223

Test 41.—Suggestion by Progressive Weights ..... [597] 231

Test 42.—Suggestion by Progressive Lines ..... [601] 235

Test 43.—Suggestion of Line-Lengths by Personal Influence ..... [607] 241

Test 44.—Suggestion by Illusion of Warmth ..... [611] 245

CHAPTER XI. TESTS OF IMAGINATION AND INVENTION

Test 45.—Ink-Blots ..... [619] 253

Test 46.—Linguistic Invention ..... [625] 259

Test 47.—Word-Building ..... [639] 273

Test 48.—Ebbinghaus' Completion-Method ..... [648] 282

Test 49.—Interpretation of Fables ..... [665] 299

CHAPTER XII. TESTS OF INTELLECTUAL EQUIPMENT

Test 50.—Size of Vocabulary ..... [673] 307

Test 51.—Range of Information ..... [682] 316

CHAPTER XII.

Note on Serial Graded Tests for Developmental Diagnosis ..... [689] 323

Appendix I. Formulas for Converting Measures (English and Metric Systems) ..... [691] 325

Appendix II. List of Abbreviations ..... [691] 325

Appendix III. List of Materials for Part II ..... [693] 327

Index of Names ..... [695] 329

Index of Subjects ..... [699] 333

INDEX OF FIGURES

FIG.		PAGE
64.	The Mirror-Drawing Test.....[488]	122
65.	The Effect of Practise During Six Trials upon the Average Speed of Mirror-Drawing (From Yoakum and Calfee).[492]	126
66.	The Effect of Long Practise upon Mirror-Drawing (From Starch).....[493]	127
67.	Percentiles for the Substitution Index for Cincinnati Working Children 14 and 15 Years Old—Sheet 1 (After Woolley and Fischer).....[510]	144
68.	Percentiles for the Substitution Index for Cincinnati Working Children 14 and 15 Years Old—Sheet 4 (After Woolley and Fischer).....[511]	145
69.	Effect of Practise Periods of Different Lengths upon Performance in the Substitution Test (From Starch).....[512]	146
70.	Jastrow's Memory Apparatus.....[525]	159
71.	Development of Memory for Digits (From Smedley)....[543]	177
72.	Percentiles of Memory for Letter Squares, Boys (Anderson).....[546]	180
73.	Percentiles of Memory for Letter Squares, Girls (Anderson).....[547]	181
74.	Memory Capacity of 12-Year-Old Pupils, by Grades (From Smedley).....[563]	197
75	Size-Weight Illusion in the Feeble-Minded (Doll).....[595]	229
76.	Guidi's Apparatus for the Warmth Illusion (Modified by Whipple).....[616]	250
77.	Percentiles of Word Building for Boys (Anderson).....[642]	276
78.	Percentiles of Word Building for Girls (Anderson).....[643]	277
79.	Percentiles for Completion Test No. 4, Ages 14 to 17 Combined (Fraser).....[660]	294

TABLE			PAGE
50.	Comparative Accuracy of Sworn and Unsworn Statements (Stern and Borst)	[397]	31
51.	Effect of Time-Interval on Report (Dallenbach)	[401]	35
52.	Dependence of Report on its Form (Stern and Borst)	[402]	36
53.	Effect of Practise on Coefficients of Report (Borst)	[405]	39
54.	Words Written in Three Minutes by Normal Children (Pyle)	[413]	47
55.	Words Uttered in Three Minutes by Epileptics (Wallin)	[414]	48
56.	Distribution of Terms in 'Uncontrolled' Association (Jastrow, Nevers, Calkins, Manchester)	[417]	51
57.	Influences that Affect 'Uncontrolled' Series of Words or Drawings (Flournoy)	[418]	52
58.	Average Distributions for the Kent-Rosanoff Test (Compiled from Eastman, Kent, Rosanoff and Strong)	[426]	60
59.	Dependence of Distribution in the Kent-Rosanoff Test on Age (Isabel Rosanoff and A. J. Rosanoff)	[427]	61
60.	Individual Responses of Children and Adults in the Kent-Rosanoff Test (Isabel Rosanoff and A. J. Rosanoff)	[428]	62
61.	Types of Associative Response in Normal and Feeble-Minded Children (Otis)	[429]	63
62.	Distribution of Responses as Conditioned by Intelligence (Isabel Rosanoff and A. J. Rosanoff)	[432]	66
63.	Performance in the Part-Whole Test (Norsworthy)	[441]	75
64.	Correct Associates Written in 60 Sec. Part-Whole Test (Pyle)	[441]	75
65.	Performance in the Genus-Species Test (Norsworthy)	[444]	78
66.	Correct Associates Written in 60 Sec. Genus-Species Test (Pyle)	[444]	78
67.	Correct Associates Written in 60 Sec. Opposites Test (Pyle)	[448]	82
68.	Correct Associates of a Possible 20. Opposites Test (Norsworthy)	[449]	83
69.	Correct Associates and Times. Opposites Test (Squire)	[449]	83
70.	Errors and Times. Opposites Test (Carpenter)	[449]	83
71.	Speed in Sec. for Correct Responses in the Analogies Test (Fraser)	[458]	92
72.	Average Scores of 29 Normal and 6 Insane Subjects in the Addition Test (Specht and Bischoff)	[474]	108
73.	Efficiency in Addition: Five 10-Minute Periods (Schulze)	[477]	111
74.	Efficiency in Addition and Multiplication within an Hour (Burgerstein)	[478]	112
75.	Additions per Pupil, with and without a Rest-Pause (Burgerstein-Schulze)	[479]	113
76.	Efficiency in Computation within a School Session (Laser)	[480]	114

TABLE PAGE

77. Effect of Practise on Speed in Mirror-Drawing. College Students (Whipple) ..... [489] 123
78. Times, in Sec., for Mirror-Drawing (Yoakum and Calfee) . [490] 124
79. Times, in Sec., Used in the Star Test by 36 College Girls, 16 College Maids and 69 Bedford Reformatory Women (Weidensall) ..... [495] 129
80. Scores in the Star Test for Three Groups of Bedford Reformatory Women (Weidensall) ..... [495] 129
81. Substitution Test. Number of Symbols Written. Form B. Group Method (Whipple) ..... [505] 139
82. Substitution Test. Speed in Seconds. Form B. Individual Method (Whipple) ..... [505] 139
83. Correct Substitutions Made in 60 Sec. Digit-Symbol Test (Pyle) ..... [505] 139
84. Correct Substitutions Made in 60 Sec. Symbol-Digit Test (Pyle) ..... [506] 140
85. Substitution Index, in Sec., Cincinnati Working Children (Woolley and Fischer) ..... [507] 141
86. Average Number of Substitutions Made by 37 White and 30 Negro Girls in a Pennsylvania Reformatory (Baldwin) . . [507] 141
87. Substitution Test. Distribution of Gains and Losses in Speed (Whipple) ..... [509] 143
88. Substitution Test. Bright and Dull Boys. Form B. Individual Method (Whipple) ..... [513] 147
89. Differences, in Sec., Index of Substitution, Form C, between Bedford Hills Reformatory Women and Cincinnati Working Girls 14 and 15 Years Old (After Weidensall) ..... [515] 149
90. Use of the 'Footrule' Method in Scoring the Memory Test (Spearman) ..... [527] 161
91. Norms of Memory Span for Digits, as Conditioned by Age (Smedley) ..... [537] 171
92. Development of Memory for Digits (Smedley) ..... [538] 172
93. Dependence of Memory Span for Auditory Digits on Age (Jacobs) ..... [538] 172
94. Dependence of Memory for Auditory Digits on Age (Ebbinghaus) ..... [538] 172
95. Memory for Letter Squares, in Relation to Age and Practise (Winch) ..... [539] 173
96. Memory for Letter Squares, Score for 10 Trials (Anderson) ..... [539] 173
97. Memory for Concrete Words, in Relation to Age and Sex (Pyle) ..... [539] 173
98. Memory for Abstract Words, in Relation to Age and Sex (Pyle) ..... [540] 174

TABLE		PAGE
99.	Average Number of Sentences Correctly Repeated (After Squire and Carpenter) [540]	174
100.	Memory for Pictures of Objects (After Squire and Carpenter) [540]	174
101.	Net Efficiency of Various Memories, in Relation to Age (Pohlmann) [542]	176
102.	Distribution of Boys' Scores in Letter Squares (After Anderson) [548]	182
103.	Distribution of Girls' Scores in Letter Squares (After Anderson) [549]	183
104.	Percentage of Accuracy in Memory for 2-Place Numbers (Schuyten) [550]	184
105.	Sex Differences in Memory Span for Digits in College Freshmen (Wissler) [550]	184
106.	Memory for 9-Term Series of Different Kinds (Lobsien) [553]	187
107.	Memory for Related and for Unrelated Words (Norsworthy) [554]	188
108.	Dependence of Memory upon Form of Presentation (Pohlmann) [556]	190
109.	Relation of Memory for Auditory Digits and Intelligence (Ebbinghaus) [560]	194
110.	Relation of Memory for Digits and School Standing (Smedley) [562]	196
111.	Memory Span for Digits in the Feeble-Minded (Johnson) [564]	198
112.	Comparative Memory Capacity of Normal and Feeble-Minded Children (Norsworthy) [564]	198
113.	Correlations with Amalgamated Results of Immediate Memory Tests (After Burt) [566]	200
114.	Correlations between Memory Tests and Other Tests (After Wyatt) [566]	200
115.	Recall of Different Members of a 7-Term Series (Binet and Henri) [566]	200
116.	Marble Statue Test, Scores by Age and Sex (Pyle) [577]	211
117.	Dutch Homestead Test. Words Written and Underlined (Whipple) [577]	211
118.	Cicero Test. Distribution of 36 College Students (Whipple) [577]	211
119.	Average Percentage of Loss in Third Reproduction (Henderson) [578]	212
120.	Story of The Fire. Scores for Epileptics by Mental Age (After Wallin) [583]	217
121.	Force of Suggestion (Gilbert) [593]	227
122.	Reactions of Feeble-Minded Children to the Size-Weight Illusion (Doll) [594]	228

TABLE		PAGE
123.	The Progressive-Weight Suggestion. 24 Cases (Binet)..[600]	234
124.	Averages for Progressive-Weight Suggestion by Age (Giroud).....[601]	235
125.	Percentage of 'Yields' to Contradictory Suggestion (Binet and Henri).....[610]	244
126.	Suggestibility to Warmth. Resistance-Coil Method (Okabe and Whipple).....[617]	251
127.	Suggestibility to Warmth as Related to Age. 187 Cases (Guidi).....[618]	252
128.	Average Number of 'Names' Given to Ink-Blots (Kirkpatrick).....[624]	258
129.	Scores of Seven Adults in Developing Sentences (Sharp).[629]	263
130.	Averages by Age and Sex, aeobmt-Test (After Anderson and Pyle).....[644]	278
131.	Averages by Age and Sex, aeirlp-Test (After Anderson and Pyle).....[645]	279
132.	Completion Text No. 4. 15-Minute Limit (Terman and Childs).....[656]	290
133.	Completion Text No. 4. 10-Minute Limit. Results for Boys (Fraser).....[657]	291
134.	Completion Text No. 4. 10-Minute Limit. Results for Girls (Fraser).....[658]	292
135.	Completion Text No. 4. 15-Minute Limit. Results for 'Citizens' of the George Junior Republic (Whipple and Fraser)..[663]	297
136.	Distribution by Percentages in Interpretation of Fables (Terman and Childs).....[670]	304
137.	Norms for Interpretation of Four Fables (Terman and Childs).....[671]	305
138.	Average Vocabulary in Relation to Scholastic Status (Kirkpatrick).....[678]	312
139.	Relation of Vocabulary to Age: Method of Terman and Childs.....[678]	312
140.	Vocabularies of 20 College Students: Terman and Childs' List (Whipple).....[679]	313
141.	Distribution of Corrected Vocabulary Index, Seventy College Students (Whipple).....[679]	313
142.	Overestimation of the Vocabulary Index. Seventy College Students (Whipple).....[680]	314
143.	Dependence of Range of Information on Academic Status (Whipple).....[684]	318
144.	Dependence of Range of Information on Academic Status (Smith).....[685]	319
145.	Dependence of Range of Information on Sex. (Whipple and Smith).....[685]	319

BOSTON MEDICAL
DEC 24 1915
LIBRARY

PART II

COMPLEX PROCESSES

CHAPTER VIII

Tests of Description and Report

The two tests which are described in this chapter have certain features in common which demarcate them, on the one hand from the tests of perception and attention of the previous chapter, and on the other hand from the memory tests of the succeeding chapter, though, in many respects, they resemble these tests.

The essential idea in both of the present tests is to determine capacity, not merely to attend and observe, or to recall what has been observed, but to put the results of this observation into linguistic form. If the observer gives his account of the experience at the time of his observation, this constitutes description; if at some time subsequent to his observation, this constitutes report.

It is evident that this giving of an account of an experience, particularly if the experience be somewhat complicated in form, is a more complex psychical process than those under discussion in the tests of attention and perception. This greater complexity makes the reduction of the observer's performance to exact quantitative terms a matter of greater difficulty, but, on the other hand, the activity called forth is more akin to that demanded in everyday life, and it is for this reason that these tests have been felt to possess a peculiar value, particularly in the study of individual differences in mental constitution and mental efficiency. Again, language occupies so strikingly prominent a place in our mental economy that tests which seek to bring out the observer's ability to cast experience into linguistic form are, on that account, well worth while. This is particu-

larly the case in the second form of test, that of the report, which, in connection with the "psychology of testimony," has of late had a prominent place in psychological research.

TEST 31

Description of an object.—The description test first came into prominence through the work of Binet, who urged that the study of individual psychology may be best advanced by resort to the experimental examination of complex, rather than of simple mental processes, and who considered the description test of special value in this connection. Binet made preliminary tests with Henri in 1893 (3), and worked at the test later by himself (1, 2). His method has been followed, though not in exact detail, by LeClere (7), Sharp (9), Monroe (8), and Cohn and Dieffenbacher (5).

Materials.—Cigarette. Cancelled 2-cent postage stamp. Lithograph, entitled "Hindoos."

The cancelled stamp was used by Monroe, the cigarette by Binet. The lithograph is substituted for the different pictures that have been used by other investigators (Binet and Henri used Neuville's "The Last Cartridge," Binet a picture representing Fontaine's "Le Laboreur et ses Enfants," Miss Sharp "The Golden Wedding" and "The Interrupted Duel"), because of the impossibility of securing these particular pictures, or of the difficulty of using them under the conditions that prevailed in the original experiments (Binet's school children were well acquainted with the fable from Fontaine, for example).

If it is desired to extend the list of materials, *E* may employ other objects used by Binet (2), such as a box of matches, a penny, a leaf, etc.

For group tests, there should be at least one picture for every 5 *S*'s, one cigarette for every 2 *S*'s, and a stamp for each *S*.

For group tests, it would be desirable to secure a set of stamps whose cancellation marks were approximately the same. The stamps should be trimmed off in such a manner as to show the full border of the stamp and a narrow margin of the paper upon which it was attached.

The lithograph is one of a series called Leutemann's Types of Nations, catalogued by E. Steiger & Co., New York. It may be purchased, like all other material cited in this book, of C. H. Stoelting Co., Chicago, Ill.

Method.—(1) For the picture-test, supply *S* with writing materials; place the lithograph upright before him, about 75 cm. distant. Instruct him: "Write a description of this picture so that one who had never seen it would know all about it." Allow 10 min.

(2) For the cigarette-test, give the following instructions, and no others: "I'm going to put on this table before you a small object. I shall leave it there under your eyes. I want you to write a description of it; not to draw it, but describe it in words. You will have about 5 min. Here is the object." If *S* is busy at the end of the allotted time, or has written but a few lines, the time may be slightly extended.

(3) For the stamp-test, proceed in a similar manner, save that *S*'s are not forbidden to draw the stamp, if they wish to. The instructions may run: "Describe this postage stamp so that a person who had never seen one would know all about it." Allow 10 min., or more if needed.

VARIATIONS OF METHOD.—The problem of assigning an appropriate title to a picture or of asking appropriate questions concerning it may be regarded as a variation of the description test. For suggestions as to this test see below, under Notes.

For young children, and, indeed, for older ones under many conditions, it is better that *E* should write from *S*'s dictation, perhaps stenographically.

TREATMENT OF DATA.—In general, the results of the description test are not intended to be submitted to exact quantitative treatment, but are to be inspected for the purpose of forming an opinion of *S*'s general mental type and capacity. The papers may, however, be treated quantitatively, by (1) counting the number of words written, or (2) counting the number of lines written. *E* may, further (3), record in general terms the readiness and ease with which *S* undertakes the description, and (4) may rate his paper as a whole, with respect to its comparative merit, on a score of 10 for a satisfactory or adequate description. (5) The description may, perhaps, be classified also with respect to its general type or character, following the classification adopted by Binet, Le Clere, and others as explained below. (6) It is possible, following Cohn and Dieffenbacher, to score descriptions more formally and precisely after the manner proposed for reports (Test 32). (7) Descriptions of the postage stamp may also be catalogued with respect to the items mentioned, as was done by Monroe.

Results.—(1) The description of an object is inadequate, because it is almost invariably *simplified, i. e.,* a considerable number of its features, even important features, are unmentioned. Thus, in one of Binet's photographs, of the 22 objects or features that were mentioned at all, only 9.4 were mentioned, on the average, in each description.

(2) This simplification or reduction in the description is the result of what might be termed a process of *selection.* Certain features are mentioned in practically all descriptions, others are mentioned only occasionally. By tabulating the number of times each feature is mentioned, one may discover some of the principles which condition this selective process. Thus, in Binet's picture of the "Laborer," the old man is mentioned 36 times, his sons 30, his bed 29, the seated woman 27, etc., until we come to relatively unimportant objects that may almost escape mention at all, *e. g.,* a stick in the hands of one of the children—only 4 times in 36 descriptions. When pictures are used, persons are more often mentioned than furniture or other details of the setting of the scene.

Similarly, in the stamp-test, tabulation indicates, according to Monroe, the following order of frequency of mention: (1) word-inscriptions, (2) color, (3) number-inscriptions, (4) portrait, (5) substance, (6) form, (7) use, (8) perforated edge, (9) size, (10) cancellation, (11) ornamentations. The item *use* declines with age: all others are mentioned more frequently as age increases.

(3) *Dependence on sex.* Monroe states that girls generally mention more items than boys, and "*seem* to surpass boys in their knowledge of the postage stamp." It is not clear, however, whether this seeming superiority is due to better observation, to greater industry or to greater zeal and conscientiousness. Cohn and Dieffenbacher similarly find the descriptions by girls more comprehensive than those by boys.

(4) *Dependence on age.* Cohn and Dieffenbacher tested school children 7 to 20 years of age with a colored picture ('Puss in Boots'). There was no clear augmentation of the range of description after 10.5 years in the case of the boys. At the age of 8, the description is predominantly an enumeration of objects,

though not a single color was mentioned by boys of that age. Esthetic and interpretative features are rarely noted before the 16th year, when a reflective element is first apparent. Increase of age is characterized by an increase in organization and systematization of the descriptions. Actual errors are relatively uncommon: the few that are met with (fidelity is 97.2 per cent. among boys, 98.7 per cent. among girls) are often verbal mistakes, the remainder true errors of apprehension.

(5) *Individual differences.* In 150 accounts of the photograph, Binet found no two alike. This wealth of individuality makes the description-test at once valuable and difficult—valuable as an indication of the variety of mental constitution, difficult as to quantitative or comparative treatment. As an extreme illustration, one may contrast the following descriptions of a postage-stamp—the first by a girl of 8, the second by a boy of 16.

(*a*) "The postage stamp has a picture in it. The postage stamp costs two cents. It says united states postage on it. The man has hair braided in back of his head. The Boarder is round. It has arms on it. The shape is square. The color is red. The man is White. You can get these to the postice [post-office] for two cents. There are lines around the boarder. The back of the stamp is white. It has nomber 2 on each side of it. The man has long hair."

(*b*) "COMMENTS ON THE ACCOMPANYING U. S. OF AMERICA 2 CENT POSTAGE STAMP.

"1. Its meaning: The Postage stamps have glorious history. In the past 57 years they have been more and more useful until now they are not only absolutely necessary, but constitute one of the great helps in the study of Geography, and one of the noblest pleasures for thousands and millions of people; Kings and Queens as well as children in the most miserable social condition.

"2. This Postage Stamp has the red color and is now next to the one penny stamps of Great Britain the most extensively used stamp used in the world. If I am not wrong its circulation in the past and present is the next largest of all others. The one penny stamp, I think has the first place.

"3. Its surroundings are very interesting. It is mounted on a piece of paper, remainder of an envelope, which fact easily indicates that it is used in the most cases for letter correspondence. I notice..... [Continues in this and the fourth paragraph a description of the stamp itself.]

"5. Some particular observations. I had 500-600 of them at home which my cousin had the kindness to send me. Of course they are of no special value, but yet they teach my little brothers the important lesson that such a little thing, like a stamp, will do all the necessary things for the transportation of a letter or other mail matter from the Atlantic to the Pacific. It is very interesting to me that with the march of civilization the great Postal system of the World has increased its actions more

and more until it is now one of the chief functions under the sun. How much this single stamp has done I cannot say, but I know that some stamps, precisely like this, have done great service to the country."

(6) *Types.* Notwithstanding this diversity, investigators have sought to classify descriptions into a limited number of types. Thus, Binet proposes four types—the descriptive, the observational, the emotional (poetic, imaginative), and the erudite—each present in varying shades and degrees.

(*a*) The *describer*, or enumerator, as one might term him, merely catalogs the features of the object before him, with little regard for their interrelations, or for the meaning of the object as a whole.

Example: "The cigarette has the general form of a cylinder, cut at one end by an inclined plane where the paper is folded. It is stuffed with a rather dark brown tobacco. The paper is striped lengthwise. The paper is somewhat bruised. The tobacco projects about 0.5 centimeter from one end."

(*b*) The *observer*, though not necessarily more intelligent or clever than the describer, places more emphasis upon the interrelations of the several features that he mentions, interprets what he sees, conjectures and indicates the significance of the object as a whole. This type is also mentioned by Mrs. Bryant in her 'description-of-a-room' test (4).

Example: "A long, white, round object, composed of a paper cylinder, about ½ or ¾ centimeter in diameter, filled with what is probably Oriental tobacco. It is about 7 centimeters long and must weigh about 6 grams [really 2 g.]. It is a badly rolled, uneven cigarette, and has been handled since it was pasted. In two places, to the right and left of the middle, the paper shows streaks as if it had been twisted. Other horizontal depressions indicate that there has been some pressure exerted upon the cigarette. I don't see the line where it has been stuck, but it must be badly fastened."

(*c*) The *emotional*, imaginative, or poetic *S* is less accurate in observation, but introduces emotion, sentiment or imaginative interpretation in his description.

Example: "It is a cigarette. It is thin, long, somewhat wrinkled. Its shape suggests a kind of elegant ease. Is it the cigarette itself or the memories that it awakes that remind me somehow of a scape-grace? The cigarette, there, all by itself on the table, makes me think of the bad student that goes off in the corner by himself to smoke. But I must write about the cigarette itself, and banish the idea of the smoker," etc.

(*d*) The *erudite S* tells what he knows, what he has been taught, or interjects bits of personal information about the object. This may indicate the presence of an unusual fund of information, or it may indicate sheer laziness, in that it is often easier to write what one knows than actually to describe from direct inspection.

Example: "We have before us here a cigarette. Let us see how it is made. In the first place, the exterior envelope is of light paper, called silk-paper. Then, inside is the tobacco. Tobacco is a product that grows almost everywhere in warm or temperate climates. The leaves of this shrub are gathered, and, after a treatment which lasts four years, are turned over to the public in the form of powder, that is, snuff, or in shreds, as in the present instance," etc.

Miss Sharp did not attempt a classification into types, but noted that *S*'s observation "may be primarily directed to the particular objects or details of the picture, to the general arrangement of the objects, that is, the composition of the picture, or to the meaning of the picture, the story which it conveys,—the details observed being such as lead up to this interpretation, or explain and apply the interpretation that is given first. The different ways in which the same picture appeals to the various individuals indicate differences in mental constitution."

The results of LeClere's test are not directly comparable with those of other investigators, because his instructions were not to describe the object (gold watch), but to "write something that comes into mind as you look at it." He distinguishes in the contributions made by 30 girls, aged 13 to 17 years, seven types, viz.: description, observation, imagination, moral reflection, erudition, pure or simple emotion, and esthetic emotion. He does not find, however, that any one of his *S*'s contributes a paper that may be classified in any one of these types, nor does any paper give evidence of a 'complete mind,' in the sense that all seven of the types are represented therein. In general, older or relatively more intelligent children write more varied or complex papers, *i. e.*, approach the theoretically 'complete' type of mental constitution.

Mention may be made here of the use of pictures in the Binet-Simon scale with the simple question: "What do you see in that picture?" or "What is that picture about?" Credit is given the child according as his replies indicate mere enumeration or a comprehension of the total meaning of the scene depicted. Mlle. Descoeudres (6) has extended the scoring of this form of description test by assigning scores to replies of different qualities and also by noting the number of ideas expressed. She gave a credit of 1 for simple enumeration, 2 if a phrase or sentence was used and 3 if the replies showed interpretation of meaning. Application of three pictures (not those used by Binet) to 14 backward children showed that rank in quality correlated distinctly with rank in quantity and also with estimated intelligence (.84, *P.E.* .02).

NOTES.—The attempt to use the description-test for classification of *S*'s into types of mental constitution is of obvious interest. The drawing, from such a classification, of inferences as to the mental make-up of the *S*'s is as obviously hazardous, for *S* may write his description in the vein that he thinks is wanted by *E*. Thus, Binet had reason to think that several *S*'s that he had classed as poetic or emotional were actually, in their everyday life, of a very matter-of-fact and unsympathetic disposition. In general, the drawing of inferences from the work of *S*'s would become safer in proportion as the descriptions were increased in number and variety, *i e.*, an *S* who wrote in an emotional vein in four descriptions of four different objects has, presumably, a real emotional constitution.

What may be regarded as a modification of the description test is the test employed by Squire (10), in which children of

various school grades were shown a series of 5 pictures and asked in each case (*a*) to supply an appropriate title to the picture, and (*b*) to ask an appropriate question about the picture. The titles proffered by the children were classed under five rubrics: "mere enumeration of objects, description of pictures, unification in terms of action of principal figures, superficial unification in terms of relation to principal object and complete comprehension evidencing imaginative insight."

With regard to the first problem, Mrs. Squire concludes: (1) No six-year-old child can be expected completely to comprehend a situation presented pictorially. (2) Neither can a seven-year-old child be expected to give an adequate title. (3) The eight-year-old children are inclined to interpret meaning in terms of action, and a few are able to give superficial titles. (4) In the ninth and tenth years the titles given are mostly descriptive, but put tersely, rather than in disjointed statements. (5) By the twelfth year the majority of the names given will pass for titles, though a large proportion still deal with superficial aspects. (6) There are many cases of complete comprehension in the thirteenth year.

With regard to the second problem, replies may be classed as failures, irrelevant, minor or essential. There are no failures after the eighth year, while the percentage of 'essential' questions rises from 6 at age 7 to 58 at age 13.

It is scarcely necessary to add that these conclusions obtain only for the particular pictures employed by Mrs. Squire, whose article should be consulted by those who seek to repeat this form of test of comprehension.

REFERENCES

(1) A. Binet, Psychologie individuelle. La description d'un objet. *AnPs*, 3: 1896 (1897), 296-332.

(2) A. Binet, L'étude expérimentale de l'intelligence. Paris, 1903. Pp. 309.

(3) A. Binet and V. Henri, La psychologie individuelle. *AnPs*, 2: 1895, (1896), 411-465.

(4) Sophie Bryant, Experiments in testing the character of school children. *J. Anthrop. Inst. of Great Britain and Ireland*, 15: 1886, 338-349.

(5) J. Cohn und J. Dieffenbacher, Untersuchungen über Geschlechts-, Alters- und Begabungs-Unterschiede bei Schülern. *Beihefte zur ZAngPs*, 2: 1911, pp. 213.

(6) Alice Descoeudres, Exploration de quelques tests d'intelligence chez des enfants anormaux et arriérés. *ArPs* (*f*), 11: 1911, 351-375.

(7) A. LeClere, Description d'un objet. *AnPs*, 4: 1897 (1898), 379-389.

(8) W. Monroe, Perception of children. *PdSe*, 11: 1904, 498-507.

(9) Stella Sharp, Individual psychology: a study in psychological method. *AmJPs*, 10: 1899, 329-391.

(10) Carrie R. Squire, Graded mental tests. *JEdPs*, 3: 1912, 363-380, etc., especially 373f.

TEST 32

Fidelity of report (*Aussage* test).—Capacity to observe, or range of observation, may be tested by methods previously described (Tests 24 and 25); native retentiveness or capacity for recall may be tested by methods such as those that are described in subsequent sections; capacity to describe what is seen may be tested as has been indicated in Test 31, but there exists a type of activity, that of reporting a previous experience, which in a way combines these several activities, in that it demands both attentive observation, retention, recall, and an ability to marshal and formulate the items of experience in a verbal report (*Aussage*). In studying the 'psychology of testimony,' interest has been developed of late in the direct examination by experimental methods of the capacity to report as such, and it has been found that reports may exhibit varying degrees of fidelity or reliability, more or less independently of the capacity that the reporters possess to observe or to retain experience; in other words, reports may contain discrepancies or inadequacies which are due, not only to misdirected attention, mal-observation and errors of memory, but also to lack of caution or of zeal for accurate statement, to scanty vocabulary, to injudicious phraseology, or, of course, to deliberate intent to mislead.[1]

Historically, the idea of subjecting capacity of report to test seems first to have been definitely proposed by Binet (3). Since

[1]It is true that no hard and fast line can be drawn between the report-test and the test of range of apprehension, or between it and the ordinary memory-test; in the main, however, range of apprehension implies a brief exposure followed by simple enumeration of the objects seen, so that what is tested is capacity to grasp or observe, rather than capacity to retain or to formulate. And the stock memory-test measures the *amount* of material that can be reproduced; in it the learning is usually by heart, and the reproduction is largely mechanical. In the report-test, the object is more complex, the time of scrutiny much longer than in the observation-test, while stress is placed as much upon quality as upon quantity of reproduction, especially upon the fidelity of reproduction as conditioned by such personal factors as timidity, cautiousness, assurance, skill in verbal formulation, etc. Again, the typical memory-test comprises a direct verbal reproduction of verbal material, while the typical *Aussage* test comprises a verbal presentation of material originally experienced as visual scenes (pictures, events, etc.), with or without some verbal features. Nevertheless, in the interrogation, the report-test does closely resemble an ordinary test of memory.

then, the study of the psychology of testimony has found its most enthusiastic and active expositor in Stern, who has written an extensive monograph (31) on the subject, and in whose periodicals (*Beiträge zur Psychologie der Aussage* and *Zeits. f. angewandte Psychologie*) most of the work of subsequent investigators has, directly or indirectly, appeared. The applicability of this line of work to many practical problems, particularly in the field of jurisprudence, is too obvious to need further comment.

GENERAL METHODOLOGY OF THE REPORT-TEST

1. *Choice of material.* Of the several types of material that have been elaborated for the study of the report, *e. g.*, the picture-test, the event-test, the rumor-test, etc., the first mentioned has many advantages for our present purposes.[2] Two types of picture-test are prescribed; the first closely patterned after that employed by Binet in his study of suggestibility in school children, the second more in accord with the stock picture-test, as developed by Stern, Borst, Wreschner, Lobsien, and others.

2. *Choice of exposure-time.* For pictures, times ranging from 5 sec. to 7 min. have been used, though 45-60 sec. is most usual. The principle which has controlled the choice of exposure-time for the two tests that follow is to select such a period as will permit an average *S* to examine each detail of the object once.

3. *Choice of time-interval.* For the sake of brevity, the instructions that follow prescribe a report directly after the exposure. If circumstances permit, *E* will find it of interest to extend the interval to several minutes, or even hours or weeks. The effect of lengthening time-interval has not as yet been satisfactorily determined.

4. *Choice of form of report.* There are two distinct forms of report, (1) the 'narrative' (*Bericht, recit*), (2) the 'interrogatory' (*Verhör* of Stern, *Prüfung* of Wreschner, *interrogatoire* of

[2]For a discussion of these advantages, of the several methods in detail, of the chief results, and for a general review of the whole field of the psychology of testimony, the reader is referred to an earlier discussion by the author (34). Suggestions for further tests will likewise be found therein.

Borst, *forçage de memoire* or *questionnaire* of Binet).[3] The narrative is a free account, delivered by *S*, either orally or in writing, without comment, question, or suggestion by *E:* the interrogatory is a series of prearranged questions; the replies to these questions constitute the 'deposition' (*Vehörsprodukt*). The constituent parts of the narrative or the deposition may be termed 'statements' or 'items.' Each form of report has its advantages and its disadvantages: both should be employed whenever possible.

5. *Choice of form of interrogatory.* An interrogatory is 'complete' when its questions cover all features of the experience exhaustively, and are propounded to all *S*'s in the same order and manner: an interrogatory is 'incomplete' when its questions are restricted to such as refer only to those items not mentioned by *S* in his narrative. The interrogatories that follow are designed to be complete, but *E* may, by appropriate selection, convert them into the incomplete type.

6. *Choice of questions.* The form of questioning very materially affects *S*'s deposition, particularly if the questions are of the type known as 'leading' or 'suggestive' questions. To some extent any question is suggestive, in so far as it implies that its recipient knows something. If we follow Stern, at least six types of questions may be framed, viz.: determinative, completely disjunctive, incompletely disjunctive, expectative, implicative, and consecutive.

A *determinative* question is one that is introduced by a pronoun or interrogative adverb, and is the least suggestive form of question, *e. g.*, "What color is the dog?"

A *completely disjunctive* question is one that forces the reporter to choose between two specified alternatives, *e. g.*, "Is there a dog in the picture?"

An *incompletely disjunctive* question is one that offers the reporter a choice between two alternatives, but does not entirely preclude a third possibility, *e. g.*, "Is the dog white or black?" In practise, for many reporters, especially for children, this form is virtually completely disjunctive, since a certain amount of independence is demanded for the choice of the third possibility, *e. g.*, for the answer "The dog is brown."

An *expectative* question is one that arouses a moderately strong sug-

[3]The terminology of the report-experiment has developed in Germany and France. I have been obliged to coin English equivalents—a task not always easy because the foreign terms have not been chosen with special care to secure consistency or to accord with legal phraseology. For this reason, the foreign equivalents are included here and elsewhere in the discussion.

gestion of the answer, *e. g.*, "Was there not a dog in the picture?" (This is the form used by Binet to induce moderate suggestion.)

An *implicative* question is one that assumes or at least implies the presence of a feature that was not really present in the experience, *e. g.*, "What color is the cat?" In practise, it is clear that a determinative question might become implicative if the reporter had completely forgotten the item to which it referred. (The implicative question was used by Binet to induce strong suggestion.)

The *consecutive* question is any form of question that is used to augment a suggeston that has been developed by previous questions.

7. *Choice of method of grading. Treatment of data.* In general, the adequacy of a report depends both upon its quantity and its quality: quantity is measured by the number of items mentioned or the number of questions answered (in absolute or in relative terms) and is referred to as the range of report (*Umfang, étendue*); quality is measured by the fidelity of the statements made, and is referred to as the accuracy of report (*Treue, fidélité*).

We have also at our command useful indications of the positiveness or degree of assurance that *S* places in his report. Besides (1) complete uncertainty ("I don't know" or "I have forgotten"), we may distinguish (2) hesitancy ("I think" or "I believe"), (3) positive statement or assurance of ordinary degree, and (4) attestation or attestable assurance, *i. e.*, the highest degree of assurance, as indicated by *S*'s willingness to take his oath that the statement is correct.

On this basis, the data may be subjected to treatment for the computation of a number of 'coefficients of report,' by the aid of the following simple formulas:

COEFFICIENTS OF REPORT[4]

Let P = number of possible items,
n = number of items reported (or replies made),
c = number of items reported with certainty (including attestation),

[4]The fourth formula is used by many writers, in place of the fifth, for accuracy of report; as here indicated, however, the indeterminate cases ("I don't know") are omitted from the denominator in computing accuracy.

Next to range and accuracy, the most important coefficient is probably warranted assurance (8th formula), as a high ratio indicates a good witness, who reports a large number of items both correctly and with assurance.

a = number of items whose correctness is attested under oath,
$n(N)$ = number of items reported in the narrative,
$n(D)$ = number of items reported in the deposition,
$n(r)$ = number of items that are rightly reported,
$c(r)$ = number of items that are certain and right,
$a(r)$ = number of items that are attested and right,
$a(w)$ = number of items that are attested and wrong

Then

(1) n = range of report, absolute (*Umfang, étendue*),
(2) n/P = range of report, relative,
(3) $n(N)/n(D)$ = spontaneity of report,
(4) $n(r)/n$ = range of knowledge (*Umfang des Wissens, étendue du savoir*),
(5) $n(r)/c$ = accuracy of report (*Treue, fidélité*),
(6) c/n = assurance (*subjective Sicherheit, assurance*),
(7) $c(r)/c$ = reliability of assurance (*Zuverlässigkeit der Sicherheit, Sicherheitsberechtigung, fidélité de la certitude*),
(8) $c(r)/n$ = warranted assurance (*Sicherheit der Person, assurance justifiée*),
(9) $c(r)/n(r)$ = assured accuracy (*Versicherte Richtigkeit, justesse certifiée*),
(10) a/n = tendency to oath or attestable assurance (*tendance au serment*),
(11) $a(r)/n$ = warranted tendency to oath (*tendance au serment véridique*),
(12) $a(w)/n$ = unwarranted tendency to oath (*tendance au faux-témoignage*),
(13) $a(r)/a$ = reliability of oath (*fidélité du serment*).

The determination of P, and hence of relative range of report, is often beset with difficulty; the most practical working rule is to rank as 'one item' any combination of features that forms a single natural working group, the details of which would escape individual observation under ordinary conditions. Again, P may be taken as the number of separate items mentioned by a competent S in describing the picture or test-object by direct observation. Or, as Hegge (18) proposes, P may be computed by adding all the specific items mentioned in the reports of any one of a number of S's. Obviously, the magnitude of P will tend to increase with the number of S's until a point is reached beyond which additional reports fail to affect it appreciably.

A similar difficulty arises in deciding what items and how many should be the subject of questions in the interrogatory. In general, the coefficients computed will have value only for a given picture or event and only when obtained by a given interrogatory, and the interrogatory must be constructed empirically, on the basis of actual preliminary trials, never *a priori*.

Although different errors unquestionably have different degrees of importance (to forget a man is more serious than to forget the color of his necktie), no satisfactory plan for arbitrarily 'weighting' different items has been devised.

The psychologically best method of grading is unquestionably to classify the data statistically according to various categories—such as persons, objects, colors, sizes, etc.—and to compute range, accuracy, assurance and the other coefficients for each category separately. This will greatly increase the labor of quantitative treatment, but it will afford valuable insight into the qualitative conditions of report that could not otherwise be secured: the several coefficients can, for comparative purposes, be united subsequently into a single series of coefficients for the person or persons under consideration.

A. REPORT-TEST WITH BINET'S CARD OF OBJECTS

MATERIAL.—Rectangular sheet of orange-yellow cardboard, 33.5 × 40.5 cm., to which are attached two photographs, a label, a button, a penny, and a postage stamp.[5] Watch.

METHOD.—Give *S* the following instructions: "I want to try an experiment with you to see how good your memory is. I am going to show you a large card with a number of things fastened on it. You will have just half a minute to look at it. Half a minute is a pretty short time, so you must look very carefully, because afterwards I shall want you to tell me what you have seen, and I shall ask you questions about many little details, and I want you to answer these questions exactly, if you can. Do you understand?"

Place the card directly before *S* in a good light. At the end of 30 sec., remove it and keep it well concealed. Direct *S* at once: "Now tell me everything you saw: describe it so clearly that if I had never seen the card I should know all about what was on it." The narrative is given orally by *S*, and recorded verbatim by *E*, without comment, query, or suggestion. Reread the report to *S*, and ask him to indicate what statements he is so sure of

[5] These objects are not exact duplicates of the Binet group, and the card is somewhat larger. The exposure-time and the questions of the interrogatory have been correspondingly modified.

that he would swear to their accuracy. Underline these statements.

Proceed next with the interrogatory. If possible, ask *S* the following questions in the order given.[6] Record his replies by number, verbatim, and underline all attested replies.

Interrogatory for the card of objects.

(1) Did you notice a *coin?*
(2) What kind of a coin is it? (What denomination?)
(3) Does it show 'heads' or 'tails?'
(4) Is it bright or dull?
(5) Is it in good condition, or scratched and marred?
(6) What is engraved on it? (What does it say?)
(7) How is it fastened to the cardboard?
(8) Did you notice a *button?*
(9) What is its shape?
(10) What is its color?
(11) Is it the same color all over?
(12) Is it made of cloth or of some other substance?
(13) How many holes are there in it?
(14) How is it fastened to the cardboard?
(15) Did you notice a *small picture* (*print*) near the top of the cardboard?
(16) What shape is it?
(17) What does it represent? (What is it about?)
(18) How many persons are there in it?
(19) What is the lady doing with her right hand?

[6] *S* may interfere with this program, either by anticipating the answers to some questions, or by committing errors, *e. g.*, describing an essentially different scene in the larger photograph; in such an event, *E* must devise other questions to follow up the cues thus given. Thus, if to Question 14, *S* replies "By a thread," ask further questions, *e. g.*, "Do the threads pass through the holes or around the whole button?" "Draw them." "What color are they," etc:

It is probably better to question *S* concerning objects that he fails to mention in his narrative, save that, naturally, if the first question in each group, "Did you notice ——?" is answered negatively, the remaining questions about that object are omitted. Many children fail spontaneously to recall one or more objects, but can nevertheless answer correctly questions about them, once the object is suggested.

(20) What is the other person doing?
(21) Where is he sitting?
(22) What is he looking at? Describe it exactly.
(23) Is the name of the picture printed on it?
(24) Did you notice *another picture?* (*A photograph?*)
(25) What shape is it?
(26) What does it represent? (What is it about?)
(27) How many persons are there in it?
(28) How are they dressed?
(29) Where are they standing?
(30) How many animals are there in the picture?
(31) Is the cart on wheels or not?
(32) Are there any words printed in the picture? What are they?
(33) What did you see in the background (in the back of the picture?)
(34) What did you see in the foreground (in the front of the picture?)
(35) Is the picture taken in summer or winter? How do you know?
(36) Did you notice a *stamp?*
(37) Is it American or foreign?
(38) How much is it worth? (What denomination?)
(39) What color is it?
(40) What is on it? (What picture or printing is on it?)
(41) On what part of the cardboard is it?
(42) Is it a new one or has it been used? (Describe the cancellation mark.)
(43) Did you notice a *label* (*sticker, paster?*)
(44) What color is it?
(45) What shape is it? (Is it perfectly rectangular? Draw it.)
(46) Is there any printing on it? What?
(47) Is there any border around the printing?
(48) How is it fastened to the cardboard?
(49) How is it placed on the cardboard—right-side up, slanting, or how?
(50) What color is the cardboard?

VARIATIONS OF METHOD.—(1) To shorten the experiment, omit the narrative and take only the deposition, but first ask *S* to name the objects seen. Record the number.

(2) Mature *S*'s may be tested in small groups, though this is not recommended. Both narrative and deposition must then be written by the *S*'s. For comparative purposes, the same procedure must be followed for all *S*'s, since oral and written reports cannot be assumed to be equivalent.

(3) To induce a moderate degree of suggestion, *E* may recast the questions of the above interrogatory into an expectative form and add others, *e. g.*, in place of No. 14: "Is not the button fastened to the cardboard by a thread?" In place of No. 30: "Isn't there a little dog besides the horse?" In place of No. 42: "Isn't the postage-stamp cancelled?" Or, for additions: "Isn't there a seventh object on the cardboard?" "Draw it." "Are there not four wheels on the cart?" etc.

(4) To induce a strong degree of suggestion, *E* may recast the questions given into an implicative form, and add others as desired, *e. g.*, in place of No. 9: "Draw the button so as to show the place where it is broken." In place of 30: "Are both horses of the same color?" In place of 42: "Describe the cancellation-mark on the stamp." In addition to 46: "What else does the label have on it besides 'Glass. Handle with care.'?" Or, in place of 21: "Is the little boy's mother putting her arm around him as he sits in her lap?" For additional questions, devise a number such as: "Is the lady's necktie dark brown or blue?" etc.

RESULTS.—(1) With regard to the number of objects spontaneously recalled, the following results indicate the outcome found by Binet[7] with 23 children 9 to 12 years of age and by an experimenter from the author's laboratory (12b) with 34 school children in the fourth grade of an Ithaca (N. Y.) public school:

Number of objects.....	6	5	4	3	Average.
French children.....	4	10	8	1	4.78
Ithaca children.....	14	10	6	4	5.00

Counting 1 for each right answer and 0.5 for each partly right

[7]For a detailed presentation of these results, see his book (3, pp. 255-329).

answer, the Ithaca children scored in their depositions (possible score = 50) an average of 30.5, with a maximum of 43.5 and a minimum of 6.5.

(2) Bearing in mind that the actual objects differed somewhat, the reader may compare Binet's results and our own with regard to the order and frequency of omission in the narrative:

Name of object.....	stamp	label	button	coin	small picture	large picture
French children....	10	9	4	3	2	0
Ithaca children.....	8	6	6	7	3	4

(3) In tests of older children with written narratives, Binet found little difference in the total number of objects mentioned, but marked differences in the wealth of details and the precision of their formulation.

(4) The objects have distinct individuality, *i. e.,* though *S* may forget the color or the value of the stamp, yet if he recalls the object at all, it is as a stamp, not, for instance, as "some square, greenish-colored thing." In other words, *S* recalls a thing, not a number of meaningless attributes.[8]

(5) *S*'s may report very precisely and with assurance objects or features of objects which are totally incorrect, *e. g.,* they may draw the thread fastening the button, and take oath as to its presence. Hence, testimony given with precision and detail and with the highest degree of assurance may be absolutely false.

(6) *S*'s may recall one feature of an object exactly, but fail entirely in their description of another feature of the same object, *e. g.,* recall that the label is red, but err as to its shape. It follows that, in testimony, a witness whose assertions are verified in many details may, nevertheless, err in his statements with regard to some other detail that happens not to be susceptible of verification.

(7) If *S* fails to mention an object in his narrative, but recalls it immediately in the interrogatory, his further characterization of it may be quite as accurate as that of other *S*'s who had recalled it spontaneously.

---

[8] In the author's study of range of visual apprehension, however, there appeared numerous cases of the character thus denied by Binet, for example, a nickel was recalled only as "something bright and round in the upper corner of the cardboard."

(8) In comparing different types of questions, Binet found 26 per cent. error for indifferent, 38 per cent. for moderately suggestive, and 61 per cent. error for strongly suggestive questions.

B. REPORT-TEST WITH A COLORED PICTURE

MATERIALS.—Set of four colored pictures: "Australians," "A Disputed Case," "Washington and Sally," and "The Orphan's Prayer."[9] Watch.

METHOD.—Give *S* instructions analogous to those in the preceding form of report-test, but without specifying the time of exposure. Expose the picture for 20 sec. Secure an oral narrative and deposition as directed above. Suggestions for interrogatories for two of the pictures follow.

Interrogatory for "A Disputed Case."

(1) How wide is the picture (horizontally)?
(2) How high is the picture (vertically)?
(3) Is there any border: if so, what color?
(4) How many persons are there in the picture?
Take the person on *your* right:
(5) Is he young, middle-aged, or old?
(6) What is his posture,—sitting, standing, or lying down?
(7) What is he doing?
(8) What is his facial expression?
(9) Is he bald or has he abundant hair?
(10) What color is his hair?
(11) Is he smooth-faced or has he a moustache or a beard?
(12) What color is his beard?

[9] All four pictures may be procured through C. H. Stoelting Co., Chicago, Ill. The "Australians" is a large lithograph, one of a series called Leutemann's Types of Nations, catalogued by E. Steiger & Co., New York City. It is recommended for use with large groups, numbering from 10 to 50 or more *S*'s. The "Hindoos" lithograph prescribed in Test 31 may be used with this for check tests, as it is of the same dimensions and of similar character.

The "Disputed Case" (No. 1235 of the Taber-Prang Art Co.'s collection) is recommended for use save for very young children or for large groups. "Washington and Sally" and "The Orphan's Prayer" (Nos. 699 and 1207, respectively, of the same collection) may be used for subsidiary and check tests.

(13) Does his moustache conceal his mouth?
(14) Does he wear eye-glasses or spectacles?
(15) Has he a hat on? What kind? What color?
(16) Where is his right hand?
(17) Where is his left hand?
(18) What color is his coat?
(19) What color is his shirt?
(20) Has he a collar on?
(21) What color is his necktie?
(22) What color is his vest?
(23) What color are his trousers?
(24) Does he wear slippers or shoes or boots?

Take the person on *your* left:[10]

(25-44) Repeat questions 5-24.
(45) What kind of light or lamp is used?
(46) Where is it placed?
(47) Where is the ink-well?
(48) Is there not a pen in it?
(49) What color is the dog?
(50) Is there a table or bench?
(51) How long is it (really)?
(52) What color is the table cloth or covering?
(53) Is the fringe of the same or of a different color?
(54) Name the objects on the table.
(55) How many chairs are there in the room?
(56) Is the rocking chair on your left or on your right?
(57) Is there an umbrella?
(58) Do you think it is jet-black or dark-blue?
(59) In what position is it?
(60) Name the objects in front of the table on the floor.
(61) Is there a satchel or dress-suit case in the room? Which?
(62) Is it open or shut?
(63) What do the pictures on the wall represent?
(64) How many windows are visible?
(65) Can you see any detail of outdoor scenery through them?

[10] If it is desired to economize time, omit questions 25 to 44.

(66) How many hats are there in the room?
(67) Describe and locate them.
(68) Can you recall the time indicated by the clock on the wall?
(69) What object is on your extreme right?
(70) Are there any books in this part of the room?
(71) What color is the wall?
(72) Where is the newspaper?
(73) How long did you see the picture?

Interrogatory for the "Australians."

(1) How many persons are there in the picture?
(2) How many animals?
(3) What kind of animals?
(4) What is the person on *your* left doing?
(5) What is the object behind him?
(6) What is the person in the middle of the picture doing?
(7) Has this person a beard or not?
(8) Is the man who is in charge of the dog holding him by a leash (guiding rope) or by taking hold directly of the scruff of his neck?
(9) What are the persons in the background doing?
(10) Do the persons in the foreground wear anything beside the loin-cloth?
(11) What color is their skin?
(12) What color is the dog?
(13) What is the most peculiar thing that you noted in the appearance of the men in the picture?
(14) What objects lie in the immediate foreground?
(15) Is there any water represented in the picture?
(16) Is the white man standing on the left or on the right?
(17) Is the sun represented in the picture as shining from your right or from your left? How do you know?
(18) How long did you see the picture?

VARIATIONS OF METHOD.—Test the effect of varying the time of exposure, of extending the time-interval between exposure and report, of repeating the report (narrative or interrogatory), without further exposure, two or more times at intervals of sev-

eral days or weeks,[11] of confronting *S* with the picture for careful criticisms of the report he has submitted. Though it is not advised as the best method, the substitution of written for oral narratives and depositions will permit an instructive class experiment.

Typical Results.—The following narrative by a college senior, a man of varied experience, mature, much traveled, and well trained, though of mediocre native ability, shows clearly the tendency of an adult *S* to describe a situation, a meaningful whole, rather than merely to enumerate details, as do many children. Indeed, the detail here is distinctly subordinated to the interpretative rendering. The narrative tells what the picture is about, rather than what it is.

"The picture, about 10x10 inches, represents a scene that would be typical of a rural justice of the peace and a man who has come to ask his advice on some subject. The Justice sits before his desk, an old manuscript before him, one hand on his head as if he had not yet given his decision. The office is filled with books and on one of them in the left of the picture rests his top-hat. The visitor seems to be troubled very much; his clothing denotes that he is of a different station in life. He has placed his carpet-bag on the floor and his hat near it, as a sign of great mental strain, which seems to increase as he awaits the decision. On the wall to the right is a double map of the world, showing, perhaps, that the Justice is a man of wisdom and a source of information to his neighbors. The room, furniture, the manner of dress would have denoted a time long before ours. The men seem to be about 65 or 70 years of age."

In his deposition, this student rendered an unusually full list of answers: the reply—"I don't know"—is given only twice (Questions 34 and 72). The range of report is, therefore, large, but the fidelity is relatively small, since all the statements that follow are erroneous ones from his report (those italicized are also attested statements):

The picture is 14x14 inches. The man on the right *is bald, wears spectacles, has his right hand on a paper*, wears a collar, *a purple tie*, black trousers, and slippers. The man on the left is thinking hard, has a troubled expression, wears a sandy moustache; he has his right hand in his pocket, *his left on his knee;* he wears a light-colored vest and brown trousers. The room is lighted by a candle which stands on the pile of books. There is a pen in the ink-well. The table is 14 feet long, has a light-colored cloth top *with fringe of a different color*. There are three chairs in the room, the rocker being at the left. *The umbrella is dark blue in color, and lies on the floor. There is a coat on the floor in front of the table;* there is a basket on the table. *The satchel is shut. One window is*

[11]See Ref. 34 for further suggestions.

visible. There is a chair at the extreme right of the picture. The wall is white. (The cuspidor and the newspaper are not recalled.)

GENERAL RESULTS OF TESTS OF REPORT.—(1) *Accuracy.* The chief single result of the *Aussage* psychology is that an errorless report is not the rule, but the exception, even when the report is made by a competent *S* under favorable conditions. Thus, in 240 reports, Miss Borst found only 2 per cent. errorless narratives and 0.5 per cent. errorless depositions. These errorless reports are commonly characterized by very small range, *i. e.,* they are reports of *S*'s who are extremely cautious and state only what they are certain of. For certain types of material, particularly estimates of time, space, number, etc., not only are erroneous reports the rule, but the most common single answer is more likely wrong than right (Dauber).

The average *S,* when no suggestive questions are employed, exhibits a *coefficient of accuracy* of approximately 75 per cent.

(2) *Range and accuracy.* There is no general relation of range to accuracy, though, for a given *S,* it is doubtless true that there is an inverse relation between these two coefficients.[12]

(3) *Range and other constants.* There is no general parallelism between range of report and other coefficients which depend upon degree of assurance.

TABLE 50

Comparative Accuracy of Sworn and Unsworn Statements (Stern and Borst)

EXPERIMENTER.....	STERN		STERN		STERN		BORST	
	Range	Errors	Range	Errors	Range	Errors	Range	Errors
Positive statements	(100)	13.6	(100)	19	(100)	23	(100)	11.0
Sworn statements	76	11	68	7	70	14	60	8.2
Unsworn statements	24	20	32	—	30	—	40	15.5
Certain statements							97.5	10.1
Uncertain statements							2.5	44.0

Note.—All figures are in per cents. The results, save those of the third and fourth columns, refer to narratives, not depositions.

[12]The reason for this lack of general relation between range and accuracy is presumably that there are two kinds of good witnesses—the one possesses good capacity of observation, recall and report, and hence exhibits a large range and a high degree of accuracy; the other is cautious, and therefore restricts his range, which may be poor at best.

(4) *Accuracy and attestation.* Generally speaking, attestation does not guarantee accuracy; on the contrary, though the number of errors is nearly twice as great in unsworn as in sworn testimony (according to Stern, 1.82 times, according to Borst, 1.89 times as great), there still remains as high as 10 per cent. error in sworn testimony. These relations are shown clearly in Table 50.

(5) *Dependence on sex.* In all of Stern's work, both in narratives and depositions, with pictures, or events, or estimations of times and distances, whether under oath or not, the reports of men have been more accurate (by from 20 to 33 per cent.), though less extended, than those of women, and a similar sex-difference has appeared in some tests of school children. This superior accuracy of boys becomes more evident when the report is difficult to make. Stern's conclusions, however, have not been confirmed by Wreschner, Breukink, or Miss Borst. Wreschner found that among adults women did better than men. Breukink found that men students reported slightly more than women, but with less accuracy, especially when colors were concerned. His men, however, proved more resistant to suggestive questions. Miss Borst, similarly, declares women to be superior to men, but an inspection of her results shows that the superiority of women consisted in the fact that they returned a larger number of correct statements, while the men did not make less accurate statements in their more limited reports. A recent and as yet unpublished investigation conducted by Boring (6) in the author's laboratory, in which groups of boys and girls and of men and women reported upon the events displayed in a moving picture leads to the conclusion that relatively little sex-difference exists between boys and girls (with a tendency in favor of the boys), whereas a quite marked and certain superiority of men over women exists among adult *S*'s.

More specifically, Borst found that in the narrative the range of men was 76 per cent., and in the deposition 83 per cent. of the range of women, while the accuracy of men in both forms of report was approximately 96 per cent. of the accuracy of women.

There is a similar discrepancy between Stern and Borst with regard to the tendency to attestation; the former found that men swore to 71 per cent. and women to 85 per cent. of their report, whereas the latter found

that men swore to 61 per cent. and women to but 59 per cent. of their report.

Boring found evidence that boys tend to exceed girls in range of report, tendency to oath and unwarranted tendency to oath, while girls undoubtedly exceed boys in reliability of oath. With adults, men apparently exceed women in range of report, and they undoubtedly exceed them in range of knowledge, assurance, warranted assurance, assured accuracy and reliability of oath. Women possess a very decidedly greater unwarranted tendency to oath. The fact that sex-differences in report are more pronounced in adults than in children accords with what we know of sex-differences in general.

(6) *Dependence on age.* Most experimenters conclude that the reports of children are in every way inferior to those of adults, that their range is smaller, their inaccuracy greater, and their warranted assurance and reliability of assurance much lower because their assurance is too great. Stern concludes that during the ages 7 to 18 the range, especially the range of knowledge, increases as much as 50 per cent., but the accuracy, save in the deposition, does not increase as rapidly (20 per cent.). This development of capacity to report is not continuous, but characterized by rapid modification at the age of puberty. Nearly all experimenters have commented upon the excessive suggestibility of children before the age of puberty. Cohn and Dieffenbacher detected improvement in fidelity up to 15 years in boys, but up to 20 in girls.

Stern has endeavored to analyze in part the development of the child's capacity to report, and has distinguished four stages: (1) the very young child enumerates only isolated objects or persons (Binet's enumerator type); (2) at about the eighth year, actions are reported more carefully; (3) during the years 9-10, attention is for the first time paid to spatial, temporal and causal relations; (4) in a still later period there appears the capacity to make a qualitative analysis of the constituent features of the objects reported. Cohn and Dieffenbacher think that there should perhaps be added a fifth period, from 16 years on, when the report shows evidence of reflective and interpretative consideration.

The question as to whether the testimony of children is so imperfect as to warrant absolute exclusion from court proceedings has given rise to much discussion. Thus, Baginsky, the German specialist in children's diseases, declares that children are the most dangerous of all witnesses and demands that their testimony be excluded wherever possible. Gross, the leading German authority on criminal law and criminal psychology, however, asserts that a healthy half-grown boy is the best possible witness for simple events, that children make different errors, but no worse ones than do adults, while, in respect to freedom from prejudice, erroneous interpretation, emotion, intoxication and the like, a child is better fitted than an adult to give an accurate report.

Lipmann contends, quite on the contrary, that the unreliability of chil-

dren's testimony is due in part to an uncritical filling out of gaps in memory, and in part to an unskillful distribution of the attention (though the child's attention is well enough concentrated on what he does report). Heindl says that children are perfectly good observers, perhaps even more objective than adults, but that they cannot translate their observations into verbal reports skillfully.

The work of Boring was specially directed toward this controversy. He found men superior to boys in all coefficients, save assurance, assured accuracy and tendency to oath, in which there was no decided difference. Women exceed girls unquestionably in both range and spontaneity of report; women display a greater tendency to oath and a greater warranted tendency to oath, but they also display a greater unwarranted tendency to oath and a lesser reliability of oath, *i. e.*, they seem to be less cautious than girls.

(7) *Dependence on intelligence.* There is no conclusive evidence upon the relation between good report and general intelligence. Winteler found no difference in range of knowledge and fidelity of report between the three most intelligent and the four least intelligent in his classes of 10-year-old boys.

(8) *Dependence on social status.* That intelligence may, however, play a positive role is suggested by the conclusions of Breukink that physicians, professors and teachers give more extended and more accurate reports than nurses and laboring men, and that the cultured group is much less open to suggestion than the uncultured and much less liable to take oath to their answers to suggestive questions.

(9) *Defectives.* The reports of defectives, paralytics, epileptics, the insane, etc., show, as one might expect, a very high degree of inaccuracy, even when the pathological condition is not seriously developed. Such persons are also highly suggestible (de Placzek). Duprée points out that the reports of such persons are peculiarly dangerous when their deficiency is latent or concealed. Gregor found that paralytics were not very bad reporters when the conditions were all favorable, but that they fell off decidedly under less favorable conditions—long time-interval, suggestion, etc.

(10) *Dependence on time-interval.* Lengthening the interval between experience and report tends, on the whole, to reduce range and accuracy, but there is nothing like the loss in efficiency shown in typical curves of forgetting for nonsense syllables and similar material; indeed, for some *S*'s the report may be

improved in some respects after several days have elapsed. Dallenbach's figures (Table 51) may be taken as fairly typical.

TABLE 51

Effect of Time-Interval on Report (Dallenbach)

INTERVAL	NARRATIVE		DEPOSITION	
	Items Recalled	Per Cent. Error	Questions Answered	Per Cent. Error
0	765	10.5	880	14.1
5 Days	735	14.3	855	18.2
15 Days	750	18.0	854	20.7
45 Days	569	22.4	801	22.4

From his earlier tests, Stern computed a fairly constant decrease of accuracy with time, amounting, on the average, to a loss of 0.33 per cent. per day over the period of three weeks which he studied; similarly, Borst computed a decrease in accuracy of 0.27 per cent. per day during a period of six days.

Though range and accuracy seem thus to suffer with the lapse of time, assurance, as shown by the number of certain and attested statements, is not, it seems, equally affected, but shows either a surprising constancy, or, if anything, a tendency to increase. From this it may be concluded that assurance and tendency to oath are due to *S*'s 'personal equation' rather than to the freshness of his memory. It would follow, of course, that warranted assurance and warranted tendency to oath decline with the lapse of time.

In explaining the improvement found in some reports after lapse of time, Schultz contends that perseveration is one of the disturbing factors in reports made shortly after the experience; in so far as perseveration is a tendency that weakens with time, there would thus be less inaccuracy from this source of error in later reports.

Jaffa asserts, more positively, that narration directly after an event by no means gives the best result; rather the memory of the event is organized and consolidated several weeks later and then affords a far more faithful picture of the event than an account after a brief interval. It seems doubtful, however, whether such a view can be accepted as a generalization, however true it may be under some conditions.

That the lapse of time occasions various and complex modifications is also indicated by the work of Cohn and Dieffenbacher, who compared direct descriptions of one colored picture (Test 31) with narratives and depositions upon another colored picture with an 8-day interval between presentation and report. Here, while there was a positive correlation between range of description and range of narration, there appeared distinct differences in the nature of the two accounts, *e. g.*, acts and interpretations are more prominent in reports than in descriptions.

(11) *Dependence on form of report.* All authorities agree that the use of the interrogatory, whether of the complete or in-

complete form, increases the range and decreases the accuracy of the report. Thus, in comparison with the narrative, the range of the interrogatory may be 50 per cent. greater, while the inaccuracy (of the incomplete interrogatory) may be as much as 550 per cent. greater. In general terms we may say that about one-tenth of the narrative is inexact, but about one-quarter of the deposition. Typical statistics are given in Table 52. Cohn and Dieffenbacher believe that reliability should always be computed from the narrative and deposition combined, because only thus can differences in the two forms of report be eliminated.

TABLE 52

Dependence of Report on its Form (Stern and Borst)

AUTHOR	RANGE		ACCURACY	
	Narrative	Deposition	Narrative	Deposition
			Per cent.	*Per cent.*
Stern	25.5	52.1	94	67.1
Borst	40.5	65.6	89	83.0

Note.—In comparing these figures, it should be remembered that Stern used an incomplete, and Borst a complete interrogatory.

According to Breukink, the use of written instead of oral reports apparently tends to increase the number of indefinite answers, but to decrease the number of erroneous answers.

(12) . *Dependence on the type of question.* The work of Stern, Lipmann, Binet and others shows that the introduction of leading or suggestive questions decidedly decreases the accuracy of report in children and may affect seriously the testimony of uncultured adults, or even of competent adults unless the conditions are favorable. Stern (33) estimates 50 per cent. error for 7-year-old children and 20 per cent. error for 18-year-old *S*'s in replies to suggestive questions. Most experimenters have found women less resistant to suggestive questions than men. Cohn and Dieffenbacher find relatively slight differences in the suggestibility of boys and girls, though the boys tend, on the whole, to take a somewhat more critical attitude. They find that

the decrease in suggestibility with age is more marked in girls than in boys, and that greater suggestibility in dull as compared with bright pupils is evident in girls, but not in boys. These investigators call attention to the fact that a suggestive question that is introduced too abruptly (so that its very form attracts attention) is apt to arouse immediate resistance.

(13) *Dependence on contents or features.* Not all the features of the original experience are reported with the same frequency or with the same accuracy. In general, we may say that persons and their acts, objects, things and spatial relations are reported with considerable accuracy (85-90 per cent.), whereas secondary features, especially quantities and colors, are reported with considerable inaccuracy (reports on color have an error of from 40 to 50 per cent.) In his subsidiary test with geometrical forms of different shapes, sizes and colors, Dallenbach found the errors most frequent with color, next with position, next with size, and least with shape, and this regardless of time-interval. Of the colors, errors were most frequent with green and least frequent with yellow tone. On the reliability of different classes of *S*'s with respect to different features, see further the tables of Cohn and Dieffenbacher (11a, pp. 86f.).

(14) *Dependence on the ideational type of the reporter.* The best reports are given by observers of a mixed ideational type, *e. g.*, acoustic-motor or visual-motor (Borst): even in a picture-test, the purely visual-minded observer is inferior, though less open to suggestion (Lobsien).

A characteristic analysis of reports, for the purpose of classifying reporters into ideational types has been given in the description-of-an-object test (No. 31), in which Binet distinguishes four types of reporter—the observer, the describer, the emotionally-minded, and the erudite. Miss Borst was unable to use this classification, however, with her *S*'s.

Another classification of reporters according to mental type was attempted by Miss Borst, who, after a preliminary tachistoscopic test, compared the reports of 'fixating' and 'fluctuating' *S*'s, and concluded that *S*'s whose attention is of the 'fixating' type have uniformly the greater warranted assurance of report. There was no relationship found with extent of report.

(15) *Qualitative analysis of errors.* Stern finds four kinds of errors in the narrative: (*a*) errors of apprehension (observation), like overlooking, misapprehending, underestimating, over-

estimating, etc.; (*b*) real errors of memory, like forgetting, filling in of gaps, gradual amplification, etc.; (*c*) errors of imagination, 'retouching' the recollection, unintentional blending of imagined experiences with the one reported, or the harmless 'playing' with the report (*Fabulieren*) often seen in children, and (*d*) errors of judgment (will), like lack of caution or self-criticism.

Schultz (29) has also attempted a qualitative analysis of the material gathered by Aall in an event test. The following are the main points upon which stress is laid:

(*a*) Whether an item is reported depends both upon the mental state at the moment and also upon the objective complex in which the item occurs. Attention is attracted by novelty and by the logical significance of the impression.

(*b*) But there is a certain 'spread' of attention such that details that are trivial and accessory may also be included with those that *S* is aiming to observe.

(*c*) The novel attracts attention, but it is also difficult to observe correctly. Optimal conditions are given when a familiar thing (easy to understand) is in an unfamiliar setting (motive of novelty).

(*d*) An event which suddenly breaks into consciousness and disturbs the set of the moment is a source of difficulty until a new adaptation for it is secured.

(*e*) "Perseveration plays an important role in the mistakes of witnesses." Its falsifying effect decreases with time, and thus reports that are separated by a time-interval from the event may be better than immediately given reports.

(*f*) There takes place a process of logical elaboration, the effect of which is to emphasize the kernel of the episode and to minimize unessential details (principle of conscious economy).

(*g*) This tendency also operates to distort reports so as to make them conform to what the witness regards as the natural course of events. Portions of the episode unperceived or not understood are filled out or rearranged in accordance with this principle. Characterizations of persons especially show this tendency.

(*h*) Many *S*'s show a distinct tendency to embellish or round out their reports into good literary form, and may thus unwittingly distort their statements.

(*i*) If the experience moves *S* emotionally, his reports are strongly colored and may suffer decided modification, particularly reports upon verbal items (quotations).

(*j*) Experiments so arranged as to cause *S* to believe that his report is serious and responsible (not a mere classroom test) produce a different conscious attitude and reveal the presence of new factors, both inciting and inhibitory; in general, the effect is to augment the value of the testimony.

(*k*) That a witness should be motivated by a desire to awaken a certain judgment upon a case need not be an undesirable condition.

(16) *The effect of repeating a report.* When *S* is called upon to make his report several times, the effect of this repetition is complex, for (1) it tends in part to establish in mind the items reported, whether they be true or false, and (2) it tends also to induce some departure in the later reports, because these are based more upon the memory of the verbal statements of the earlier reports than upon the original experience itself, *i. e.,* the later reports undergo distortion on account of the flexibility of verbal expression.

(17) *The effect of practise: educability.* On the basis of Miss Borst's work (Table 53), it would appear that simple prac-

TABLE 53

Effect of Practise upon Coefficients of Report (Narrative) (Borst)

NUMBER OF REPORT (TEST)	I	II	III	IV	V
Range	39.0	39.0	42.3	40.3	42.0
Accuracy	86.6	87.7	92.9	88.2	90.0
Assurance	96.6	96.4	97.8	97.9	98.6
Warranted assurance	84.0	87.0	91.0	88.0	89.0
Reliability of assurance	87.5	89.4	92.6	89.8	90.3
Assured accuracy	97.0	98.0	98.4	98.6	99.2
Tendency to oath	43.0	59.8	62.8	61.9	72.1
Warranted tendency to oath	40.2	53.2	58.5	57.5	66.5
Unwarranted tendency to oath	2.8	6.6	4.3	4.4	5.6
Reliability of oath	93.0	88.8	92.5	93.0	91.7

Note.—The effect of practise in these tests is somewhat obscured by the fact that the first and third tests were made after a 3-day, the others after a 9-day interval.

tise, without special coaching or conscious effort to improve, facilitates the report. In her work it will be noted that the tendency to oath and warranted tendency to oath are both particularly improved, while there is appreciable improvement in the other coefficients, save assurance and assured accuracy. On the other hand, some doubt is cast upon generalizations from Miss Borst's work by the reports made by Baade and Lipmann for the Commission of the Institute for Applied Psychology appointed especially to investigate this problem of the educability

of report. Baade shows that with regard to reports made upon verbal statements (quotations) the *S*'s (196 girls, aged 12-13 years) showed no demonstrable improvement, either as a result of the threefold repetition of each experiment (physical laboratory demonstrations) or as a result of the succession of three different experiments. There was an influence of earlier upon later experiments, but this influence was sometimes favorable and sometimes unfavorable. Lipmann, who scored the estimates of duration and size, found, on the whole, some improvement in these estimates due to the succession of experiments, but only a very slight improvement due to the repetition of given experiments.

Other experimenters have reported results more nearly in accord with Miss Borst's conclusions. Breukink, for instance, found that if *S*'s are allowed to see the picture after reporting, the practise increases fidelity of report, especially in the deposition and in resistance to suggestive questions. Again, the very interesting *Methode der Entscheidungs- und Bestimungsfragen* (questions in form of: "Do you know thus and so?" and "What is thus and so?" respectively) has led Franken to declare that such training as this method induces, causes an improved cautiousness in asserting positive knowledge.

Other experiments by Marie Dürr-Borst (1906) indicate that improvement in the capacity of children may be best secured by appeal to zeal, interest, enthusiasm and desire for improvement, whereas more formal training of an intellectual type—suggestions for systematic observation, specific training in sense-perception, etc.—is much less effective.

REFERENCES

A. The most important single source is Stern's *Beiträge zur Psychologie der Aussage*, Leipzig, 1903-6. Lack of space forbids the itemizing of the numerous titles: besides extended reviews, communications, reports of lectures, etc., this periodical contains important articles by Stern, Jaffa, Cramer, Lobsien, Lipmann, Borst, Bogdanoff, Rodenwaldt, Oppenheim, Kosog, Wendriner, Günther, Gottschalk, and others.

B. The following are other important references. See especially Nos. 11, 21, 32, 34 and 35 for bibliographies and general reviews. The new literature is summarized annually in *PsBu*.

(1) W. Baade, Aussage über physikalische Demonstrationen. (Mit besonderer Berücksichtigung der Frage der Erziehbarkeit der Aussage.)

1 Abh. Die Methodik der Versuche und die Inhalte der Textaussagen. *ZAngPs*, 4: 1911, 189-311.

(2) A. Baginsky, Die Kinderaussage vor Gericht. Berlin, 1910. Pp. 41.

(3) A. Binet, La suggestibilité. Paris, 1900. Pp. 391.

(4) A. Binet, La science du temoignage. *AnPs*, 11: 1904 (1905), 128-137.

(5) A. Binet, Psychologie individuelle. La description d'un objet. *AnPs*, 3: 1896 (1897), 296-332.

(6) E. G. Boring, Capacity to report upon moving pictures as conditioned by age and sex. To appear probably in *PsRev.*

(7) Marie Borst. Recherches experimentelles sur l'éducabilité et la fidélité du temoignage. *ArPs(f)*, 3: 1904, 233-314.

(8) Marie Dürr-Borst, Die Erziehung der Aussage und Anschauung des Schulkindes. *EPd*, 3: 1906, 1-30.

(9) M. Borst et E. Claparède, La fidélité et l'éducabilité du témoignage. *Arch. des sciences physiques et naturelles*, April 7, 1904.

(10) H. Breukink, Ueber die Erziehbarkeit der Aussage. *ZAngPs*, 3: 1909, 32-87.

(11) E. Claparède, Psychologie du temoignage. (General review.) *ArPs(f)*, 9: 1910, 228-232.

(11a) J. Cohn und J. Dieffenbacher, Untersuchungen über Geschlechts-, Alters- und Begabungs-Unterschiede bei Schülern. *Beihefte zur ZAngPs*, 2: 1911. Pp. 213.

(12) K. M. Dallenbach. (*a*) The relation of memory error to time-interval. *PsR*, 20: 1913, 323-337. (*b*) The effect of practise upon visual apprehension in school children. *JEdPs*, 5: 1914, 321-334, 387-404.

(13) J. Dauber, Die Gleichformigkeit des psychischen Geschehens und die Zeugenaussagen. *FsPs*, 1: 1912, 83-131.

(14) E. Duprée, Le temoignage: étude psychologique et médico-legale. *Rev. d. deux Mondes*, 55: 1910, 343-370.

(15) A. Franken, (*a*) Ueber die Erziehbarkeit der Erinnerrungsaussage bei Schulkindern. *ZPdPs*, 12: 1911, 635-642. (*b*) Aussageversuche nach der Methode der Entscheidungs- und Bestimmungsfrage bei Erwachsenen und Kindern. *ZAngPs*, 6: 1912, 174-253.

(16) H. B. Gerland, Zur Frage der Zeugenaussage. *Arkr*, 39: 1910, 116-119.

(17) H. Gross, Zur Frage der Zeugenaussage. *ArKr*, 36: 1910, 372-382.

(18) T. Hegge, Zur Frage der Bewertung von Aussagen bei Bildversuchen. *ZAngPs*, 6: 1912, 51-59.

(19) R. Heindl, Die Zuverlässigkeit von Signalamentsaussage bei Schulkindern. *ArKr*, 43: 1909, 109-132.

(20) S. Jaffa, Ein psychologisches Experiment im kriminalistischen Seminar der Universität Berlin. *Beiträge zur Psych. der Aussage*, 1: 1903, 79-99.

(21) O. Lipmann, Neuere Arbeiten zur Psychologie der Aussage: Sammelbericht. *JPsN*, 3: 1904, 245-249.

(22) O. Lipmann, Die Wirkung der Suggestivfragen. *ZPdPs*, 8: 1906, 89-96.

(23) O. Lipmann, Die Wirkung v. Suggestivfragen. *ZAngPs*, 1: 1907-1908, 44-92, 382-415, 504-546; 2: 1908, 198-242. Also published separately, Leipzig, 1909.

(24) O. Lipmann, Pedagogical psychology of report. *JEdPs*, 2: 1911, 253-261.

(25) M. Lobsien, Ueber Psych. der Aussage. *ZPdPs*, 6: 1904, 161-209.

(26) L. Maurer, Beobachtungen über das Anschauungsvermögen der Kinder. *ZPdPs*, 5: 1903, 62-85.

(27) H. Münsterberg, On the witness stand. N. Y., 1908. Pp. 265. (The several sections of this book have also appeared in magazine form, chiefly in McClure's Magazine.)

(28) de Placzek, Experimentelle Untersuchungen über die Zeugenaussagen Schwachsinniger. *ArKr*, 18: 1904, 22-63.

(29) G. Schultz, Zur Aussagepsychologie: Prinzipielle Erörterungen im Anschluss an ein zweifaches Experiment. *ZAngPs*, 7: 1913, 547-574.

(30) C. und W. Stern, Erinnerung, Aussage und Lüge in der ersten Kindheit. Leipzig, 1909. Pp. 160.

(31) W. Stern, Zur Psychologie der Aussage. Experimentelle Untersuchungen über Erinnerungstreue. *Zeits. f. d. ges. Strafrechtwissenschaft*, 22: 1902. (Also published separately, Berlin, 1902.)

(32) W. Stern, (*a*) Literatur zur Psychologie der Aussage. *ZAngPs*, 1: 1907-8, 429-450. (*b*) Bibliographie zur Psychologie der Aussage, 1908-1910. *ZAngPs*, 4: 1911, 378-381. (*c*) Psychologie der Aussage, 1911-1913. *ZAngPs*, 7: 1913, 577-596.

(33) W. Stern, Abstracts of lectures on the psychology of testimony. *AmJPs*, 21: 1910, 270-275.

(34) G. M. Whipple, The observer as reporter: a survey of the 'psychology of testimony.' *PsBu*, 6: 1909, 153-170. (Also annual summaries of literature on the topic, *PsBu*, 1910 on.)

(35) J. H. Wigmore, Professor Münsterberg and the psychology of evidence. *Illinois Law Review*, 3: Feb., 1909, 399-445 (with bibliography of 127 titles).

(36) A. Wreschner, Zur Psych. d. Aussage. *ArGesPs*, 1: 1905, 148-183.

CHAPTER IX

Tests of Association, Learning, and Memory

A generation ago, the members of the 'English School' of psychologists exalted 'association' as a fundamental principle or law of mind comparable in its scope and importance with the law of gravitation in the material world. Whether this extreme position be held or not, it must be admitted that the more complex phases of mental activity are more readily understood if certain basic conditions of mental elaboration are posited, particularly the conditions: attention, retention, and association. Disregarding the first of these, which we have already discussed, we find in retention the *sine qua non* of the development of human mental activity, and we find constantly at work in the conscious life of the organism a tendency for the establishment of connections between its concurrent and its successive psychophysical activities. In so far as the conscious organism acquires new capacities for response, there must be retention and organization. Learning, retaining, recalling, associating, these are terms obviously descriptive of a series of related activities, and on this account, tests which deal with them are here assembled.

Association and memory, taken together, have undoubtedly been the occasion of more numerous and more elaborate experimental investigations than any other phase of mental life. Learning, in the narrower sense, has, perhaps, received somewhat less attention, though of late the importance of its application to pedagogical problems has stimulated work upon it.

The experimental study of associative activity can be, and has been, undertaken for quite varied purposes, *e. g.*, to examine the time relations of mental phenomena, to study individual differences in thought-processes, as conditioned by age, sex, training, physical condition, and the like, to analyze the diurnal curve of psychophysical efficiency (as in Kraepelin's use of computation), to diagnose mental content, and even to reveal obscure

mental tendencies and motives or intentionally withheld information (diagnostic association tests). Space forbids the exploitation of all the tests that have been developed in these fields, but a study of the more common tests of learning, association and memory that have been selected for treatment here as being most applicable to the experimental study of school children will serve to indicate the lines along which variant methods may be developed and employed.

The earlier tests in this chapter investigate the nature and efficiency of those associative connections that the subject has already established at the time of the test, either when the associative processes are allowed free rein (uncontrolled association) or when they are placed under certain restrictions (controlled association). The tests of learning that follow investigate the subject's capacity to establish new associative connections, under relatively novel conditions. The memory tests, in a somewhat different way, investigate his retentive capacity or his ability to reproduce a series of symbols or a series of related ideas. The classification of tests of association, learning and memory is, of course, somewhat rough; it is difficult to draw sharp distinctions between each type or to delimit precisely the mental processes that are brought into operation, as is illustrated, for example, in the obvious overlapping of tests of memory, of memory-span, of report, of range of attention and range of apprehension.

TEST 33

Uncontrolled association—continuous method.—The essence of this test is the requirement to write or pronounce an extended series of words not in the form of sentences. Our interest lies, first, in the difference of facility exhibited by different *S*'s in the production of such a series of terms; secondly, in the nature of the terms given by *S*'s of different sex, age, or social condition; and thirdly, in the nature of the mental processes underlying the word-naming process.

Cattell and Bryant (4) make brief mention of the test; Jastrow (6, 7), and later Miss Nevers (10), Miss Calkins (3), Miss

Tanner (11) and Miss Manchester (8), employed it for the study of the community of ideas of men and women, Flournoy (5) for the study of the effect of environment, present and immediately past, upon the course of association, and Binet (1) for the study of individual differences in intellectual processes. In a modified form (test of 60 words in 3 min.) it appears in the Binet-Simon Scale (Ch. XIII).

MATERIALS.—Stop-watch. Blank forms containing numbered spaces for 100 words. [The seconds-clock.]

METHOD.—Give *S* these instructions: "When I say 'now,' I want you to start in with some word, any one you like, and keep on saying words *as fast as you can* until you have given a hundred different words. You may give any words you like, but they must not be in sentences. I will tell you when to stop." *E* starts the stop-watch at the command 'now' and writes on the prepared form the words spoken by *S*. With mature *S*'s, it may be possible to get nothing more than scant abbreviations for the more rapid portions of the series, but these may be filled out subsequently. The points at which *S* makes distinct pauses may be noted on the form. At the 100th word, stop the watch and record the time. If time permits, and *S* can do so, it is advisable at once to go over his series, not only to fill out the list of terms, but also to make marginal notes of all the intermediate links and subsidiary associative processes that he can recall.

VARIATIONS OF METHOD.—(1) For group tests, *E* should provide each *S* with a blank. He may allow 3 min. for writing, and rate speed in terms of number of words written (method followed by Pyle, 10), or each *S* may record his own time for writing 100 words by the aid of the seconds-clock for group tests. The latter procedure is recommended rather than the former. When *S* does the writing, the method resembles that of the users of it discussed below, but the standard method of oral naming is best.

(2) Instruct *S* to keep his eyes closed during the test. This variant is to be preferred for individual testing, at least with adults; its effect is commonly to reduce the speed of naming and to lessen the number of terms suggested by objects visible in the room where the test is administered.

(3) Vary the test by demanding short lists, say of 20 words each, referring to the several categories indicated in the table which follows, *e. g.*, "Name words pertaining to clothing." "Name abstract terms." "Name adjectives," etc. Note the time needed for each such list.

(4) *E* may omit the instruction to write or to speak as rapidly as possible, and allow *S* to work at his leisure. This method, which was followed by Miss Nevers, is perhaps more satisfactory for the subsequent qualitative report upon the series, but deprives the test of whatever quantitative merits it possesses, besides tending to yield results of a distinctly different nature that are not comparable with those otherwise obtained.

(5) When working with younger *S*'s, *E* may with advantage limit the length of the series. Thus, Flournoy demanded but 10 words, while Binet recorded the time for three series of 20 words each, and occupied the intervals in reviewing with *S* the terms of the preceding series. This method is less fatiguing, and enables immature *S*'s to give a more satisfactory account of their associative connections, but it does not test *S*'s capacity as rigorously as the longer list.

(6) *E* may secure a very limited measure of uniformity in the earlier portion of the series by starting all *S*'s from the same word. For this, the words *quick* and *play* are recommended. Here it is of interest to observe the lines of divergence in association taken by different *S*'s.

(7) Another variation is that of Flournoy, who, in addition to the word test, gave 45 *S*'s instructions to make 10 drawings of any sort.

Treatment of Data.—In the standard form of test, *S*'s speed is indicated directly by his time for naming 100 words. In the group test, it is customary, similarly, to rate *S*'s speed in terms of words written in 3 min. It is not possible, however, to regard the times obtained from these two forms of the test as interchangeable, since the second form includes writing and this, as is demonstrated below, tends, even in the case of mature *S*'s, to slow the rate of performance. In so far, too, as *S*'s differ in their speed of writing, this fact enters as an unavoidable disturbing factor in the group test.

For qualitative comparison of the lists, *E* may, by inspection, supplemented by *S*'s explanation, catalog the words, either in the 7 categories used by Binet, or in the 25 categories used by Jastrow, Miss Nevers and Miss Manchester. Both classifications are embodied in the results below.

RESULTS.—(1) Some idea of the relation between performance in the group test (words *written* in 3 min.) and *age and sex in normal S's* may be secured from the averages published by Pyle for a limited number of cases and under less precise instructions than those above recommended. These results are set forth in Table 54, where it will be observed that on the whole the number of words increases with age year by year, and that girls at nearly every age somewhat excel boys in their scores.

TABLE 54

Words Written in Three Minutes by Normal Children (Pyle)

SEX	AGE	8	9	10	11	12	13	14	15	16	17	18	ADULT
Male	Cases	33	60	66	66	77	80	57	38	36	16	21	64
Male	Aver.	23.0	26.9	29.7	33.3	34.2	33.9	33.3	40.0	33.3	42.8	48.9	42.2
Male	Av. Dev.	7.5	7.6	9.0	11.4	10.9	14.6	13.2	14.8	14.6	12.3	16.6	13.8
Fem.	Cases	37	82	88	65	90	66	61	46	46	38	29	86
Fem.	Aver.	23.7	31.0	32.2	36.8	36.6	38.3	39.1	40.2	40.9	41.6	47.1	38.3
Fem.	Av. Dev.	8.2	8.9	10.8	12.1	15.4	16.8	12.9	13.8	14.1	14.0	13.9	13.1

(2) In tests of college students the average time for writing 100 words ranges between 5 and 6 min. Jastrow reports an average of 130 sec. for oral and 308 sec. for written lists of this length. Since writing an equal number of words from dictation took 212 sec., he concludes that about 1.14 sec. was used, on the average, in thinking the association between one word and the next.

(3) Table 55, derived from Wallin's studies of *mentally defective epileptics* (12), shows that the test of uncontrolled association (here the number of words *spoken* in 3 min. under certain special instructions necessitated by the nature of the *S*'s) is of some value for mental classification, since the average re-

sults show a steady increase with increase in mental age (Binet-Simon diagnosis) when due allowance is made for the small number of cases tested in certain ages.

TABLE 55

Words Uttered in Three Minutes by Epileptics (Wallin)

Binet-Simon Age------	VI	VII	VIII	IX	X	XI	XII	XIII
Average Words Spoken-	16.0	25.5	21.5	33.4	43.6	51.3	59.9	65.0

(4) Inspection of the lists printed both by Jastrow and by Binet shows that *S*'s follow what might be termed a *series of themes:* a number of terms are written, all of which cluster about a common central idea; through one of these terms access is given to a new central idea, which in turn becomes a theme for the next series of terms. Thus, in the series *hand, face, lip, chest, knees, calf, cow, horse, pig,* etc., the transition from the parts-of-the-body theme to the animal theme is effected by the common term *calf.*

(5) In some *S*'s, the controlling theme is an *auditory sequence,* which occasions long series of rimed or alliterative terms, *e. g., run, pun, fun,* etc., or *hen, hand, head, harp,* etc.

(6) In this test, the *most common words, i. e.,* those most easily got at, or those that lie, as it were, on the surface, are given first. After these are delivered, the task grows more difficult; deeper and more remote-lying terms must be actively sought for. Closely related to this is the fact that, at least in the lists of younger *S*'s, practically all the terms are nouns.[1] This is particularly the case in the short series conducted by Binet, so that, as he remarks, the test, as he conducted it, is virtually equivalent to a request to write 20 common nouns.

(7) In view of the vast number of words available, it is at first surprising to note the *degree of community* present in lists

[1]Children often interpret the instructions to mean that only nouns are wanted. If *E* stops to explain that other parts of speech are permissible, the result is sometimes more confusing yet, as they may then seek to name some of every part of speech.

of 100 terms given by a limited number of persons. Thus, Jastrow found that in 50 lists (5000 words), only 2024 words were different, only 1266 words occurred but once, while the 100 most frequent words made up three-tenths of the whole number.

These most frequent words are, as has just been said, names of common objects: in Jastrow's 50 lists, the following were the most frequently used words: *book* (40), *horse* (37), *girl* (35), *man* (34), *boy* (33), *table* (30); then follow *chair, tree, cow, paper, dress,* etc., in somewhat lesser frequency

(8) For the *classification* of the words given by 20 12-year-old pupils, Binet found seven categories adequate, viz.: (*a*) names of objects in the room where the test was held, (*b*) parts of the person or clothes, (*c*) objects or persons in the school, (*d*) objects recalled from the home, (*e*) objects seen in the streets (horse, tree), (*f*) objects seen in fields or on country excursions, (*g*) unclassified nouns. Here there is no place for abstract terms, many of which were found in series given by American pupils in Jastrow's tests. Jastrow's own classification is indicated in Table 48, where it will be seen that his 25 categories are much more elaborate and extended than those employed by Binet.

(9) *Dependence on sex.* The question as to sex difference in spontaneous trains of ideas such as are evoked in this test has been answered differently by the tests conducted at different institutions. The comparison of Wisconsin men and Wisconsin women was made by Jastrow, the 1894 test of Wellesley women by Miss Nevers and with no instruction as to speed, the 1896 test of Wellesley women by Miss Calkins but with the same instructions as those of Jastrow, the test of 75 men and 75 women at the University of California in 1905 by Miss Manchester after Jastrow's method. The categories of particular interest are those printed in italics. Jastrow's results in this and other tests led him to believe that "women repeat one another's words much more than the men." He found that "the class to which women contribute most largely is that of articles of dress, one word in every eleven belonging to this class. The inference from this that dress is the predominant category of the feminine (or of the

BOSTON MEDICAL
DEC 24 1915
LIBRARY

privy feminine) mind is valid with proper reservations." Since the women exceed the men in the enumeration also of foods, amusements, arts, and educational matters, but fall below them in naming implements and utensils, professions, and especially in abstract terms, Jastrow concludes, "that the feminine traits revealed in this study are an attention to the immediate surroundings, to the finished product, to the ornamental, the individual, and the concrete, while the masculine preference is for the more remote, the constructive, the useful, the general, and the abstract" (6; pp. 564-5). Most of these conclusions are flatly opposed by the Wellesley results of 1894, but the employment of Jastrows' methods in the 1896 test produced less marked divergencies. It is particularly to be noted that writing at a faster rate (1896 test) caused a marked decrease in the number of abstract terms, and brought the terms relating to 'interior furnishings' up even beyond those of the Wisconsin women; on the other hand, the frequency of terms for 'wearing apparel' was not affected by this change in method.

The three sets of 25 lists each (25 men and 25 women) procured by Miss Manchester at California show complete agreement with Jastrow's results in the following aspects: men lead in naming (1) verbs, (2) implements and utensils, (3) occupations; women lead in naming (1) wearing apparel, (2) buildings and building materials,[2] (3) interior furnishings, (4) educational terms, (5) arts, and (6) amusements. Miss Manchester generalizes these differences as follows: (1) "The dynamic aspect of objects is more attractive to men, while the static or completed aspect appeals more to women." (2) "Time as a factor enters more largely into the surface ideas of men; space is more often a prominent feature of the surface ideas of women." (3) "Men are interested in far-reaching relations existing between things; women give more attention to the minute analysis of things themselves." (4) "The range of the surface ideas of men, as a group, is slightly greater than that of women."

---

[2]In explanation of this seemingly unusual superiority of the women it should be said that the things named are not distinctive building materials or operations, like mortar, cement, mortising, etc., but such common terms as *floor*, *door*, *gate*, *church*, etc.

TABLE 56

Distribution of Terms in 'Uncontrolled' Association (Jastrow, Nevers, Calkins, Manchester)

(*Each column represents 25 lists of 100 words each. Those from California are based upon 75 lists reduced to the same basis.*)

CATEGORIES	WISCONSIN MEN	WISCONSIN WOMEN	CALIFORNIA MEN	CALIFORNIA WOMEN	WELLESLEY WOMEN, 1896	WELLESLEY WOMEN, 1894
1. Animal kingdom	254	178	214	187	146	223
2. Wearing apparel and fabrics	129	224	82	118	97	96
3. Proper names	194	153	84	92	81	141
4. Verbs	197	134	302	258	279	114
5. Implements and utensils	169	121	115	82	139	132
6. Interior furnishings	89	190	90	119	212	84
7. Adjectives	177	102	208	266	300	234
8. Foods	53	179	81	78	88	56
9. Vegetable kingdom	121	110	83	90	101	91
10. Abstract terms	131	97	113	101	101	280
11. Buildings and building materials	105	117	121	140	86	106
12. Parts of body	101	105	91	62	66	34
13. Miscellaneous	91	97	197	180	123	162
14. Geographical and landscape features	97	80	102	114	70	142
15. Mineral kingdom	74	96	96	58	30	54
16. Meteorological and astronomical	85	76	86	87	109	26
17. Stationery	60	86	58	54	69	26
18. Occupations and callings	71	47	60	35	24	33
19. Conveyances	62	52	44	50	19	79
20. Educational	34	76	59	74	102	167
21. Other parts of speech	96	5	108	103	164	41
22. Arts	33	61	59	79	17	44
23. Amusements	30	53	25	45	17	102
24. Mercantile terms	30	29	13	14	18	15
25. Kinship	17	32	9	12	42	18

Burt and Moore repeated Jastrow's test both with children and adults of both sexes, with results that roughly confirm Jastrow's. "The females are more personal and subjective in their interests; the males are more impersonal and objective. . . .

Women alter their themes and topics far more frequently than men; men, on the other hand, show a greater variety of associative connections between one idea and another within the same theme. The course of ideas is also more frequently disturbed in women by the various signs of 'complexes' (systems of associated ideas characterized by strong emotional colorings)."

These discrepancies raise the issue, as Miss Tanner has pointed out, whether this test can be expected to reveal fundamental native differences in mental constitution of the two sexes, or whether it reveals merely acquired traits, social traditions, individual habits, educational and other environmental influences. The lists written by college students might be expected, for example, to be considerably affected by their recent occupations, courses of study pursued at the time, etc.

The more direct comparison of the speed of the two sexes in naming terms of different kinds which we have suggested (Variation of Method, 3) does not appear to have been attempted by any of these investigators.

(10) This *influence of environment* upon the lists of associations is indicated particularly in Flournoy's brief tests (10 words and 10 drawings), the results of which are summarized in Table 57.

TABLE 57

Influences that Affect 'Uncontrolled' Series of Words or Drawings (Flournoy)

	DRAWINGS	WORDS
	Per cent.	*Per cent.*
Traced to present surroundings	13.8	29.0
Traced to the immediate past	1.9	8.2
Due to the milieu	15.7	37.2
Traced to recent personal experiences	2.4	3.9
Traced to personal habits	39.2	9.2
Expressing individuality	41.6	13.1
Unexplained	42.7	49.7

REFERENCES

(1) A. Binet, L'étude expérimentale de l'intelligence. Paris, 1903. Pp. 309. Especially chs. ii to iv.
(2) C. Burt and R. C. Moore, The mental differences between the sexes. *JEPd*, 1: 1912, 273-284, 355-388.
(3) Mary W. Calkins, Community of ideas of men and women. *PsR*, 3: 1896, 426-430.
(4) J. McK. Cattell and Sophie Bryant, Mental association investigated by experiment. *Mind*, 14: 1889, 230-250.
(5) Th. Flournoy, De l'action du milieu sur l'ideation. *AnPs*, 1: 1894 (1895), 180-190.
(6) J. Jastrow, A study in mental statistics. *New Review*, 5: 1891, 559-568.
(7) J. Jastrow, Community of ideas of men and women. *PsR*, 3: 1896, 68-71, 430-1.
(8) Genevieve S. Manchester, Experiments on the unreflective ideas of men and women. *PsR*, 12: 1905, 50-66.
(9) Cordelia Nevers, Dr. Jastrow on community of ideas of men and women. *PsR*, 2: 1895, 363-7.
(10) W. H. Pyle, The examination of school children. New York, 1913. Pp. 70, especially 24-27.
(11) Amy Tanner, The community of ideas of men and women. *PsR*, 3: 1896, 548-550.
(12) J. E. W. Wallin, Experimental studies of mental defectives. *EdPsMon*, No. 7, Baltimore, 1912. Pp. 155.

TEST 33A

Uncontrolled association—discrete method (Kent-Rosanoff test).—This test resembles the preceding one in that it deals with association of the free, unrestricted or uncontrolled type, but it differs from it in that *S* is called upon to respond with a single term only to each of a series of words presented by *E*. This form of response has been, of course, the object of an extraordinary amount of investigation, particularly with reference to its time-relations. But in the special arrangement of the test developed by Kent and Rosanoff no attempt is made to measure the time-relations, and the search for devices for the logical classification of the responses (a decidedly prominent feature of many laboratory and clinical studies in association) is limited to a simple empirical sorting of them into 'common,' 'doubtful,' and 'individual' responses, on the basis of prepared frequency tables.

Tabulated lists of the frequency with which different responses are made to the stimuli presented in association tests

were constructed by Cattell and Bryant (4) as early as 1889, and since then have been developed to some extent by Gertrud Saling (20) in 1908, and by Reinhold (16) in 1910, while Bovet (1) has outlined several methods by which such tables might be handled in figuring a 'coefficient of banality.' Nevertheless, the frequency tables published by Grace Kent and A. J. Rosanoff (10) in 1910, taken in conjunction with the supplementary reports upon their applicability made by Rosanoff with the assistance of Eastman (5) in 1912 and of Isabel Rosanoff (18) in 1913 and the recent study by Miss Otis (15), constitute so important and well-standardized a development of the idea of measuring commonplaceness, or normality of response by means of empirical tables of distribution, as to warrant the introduction of their test as a special and specific method of testing association.

Whether the Kent-Rosanoff test merits the rather extravagant encomiums that have been awarded it by some writers[1] appears to me extremely doubtful; it has certainly discarded whatever advantages might be secured by resort to introspection and to the making of time measurements; it sets up an arbitrary standard of normality, valid at best only in the gross and when the test is conducted by certain fixed and probably far from ideal conditions for exploring individuality in mental connections.

Materials.—Prepared forms comprising a printed list of 100 stimulus words[2] with spaces for recording responses, their times and their indexes. The Kent-Rosanoff frequency tables. [Stop-watch.]

---

[1]Woodworth and Wells (25), for example, talk of the free association test as having achieved, and being likely to retain, a place "in the foremost rank among the methods of individual psychology," and assert that the form of it developed by Kent and Rosanoff gives "perhaps the best objective correlate of temperament at present to hand," and that it has "established a definite standard of normality"—statements that are hard to understand in the light of the results established to date.

[2]Sixty-six of these terms are taken from the series published by Sommer in his *Diagnostik der Geisteskrankheiten;* the remainder have been selected on the basis of preliminary experimentation in such a manner as to cover a variety of situations without being especially liable to call up personal experiences.

METHOD.—Seat *S* in a room free from distracting influences and with his back to *E*. Instruct him as follows: "I am going to read to you, one at a time, a series of 100 words. Just before each word I shall call out 'ready.' As soon as you hear the word that follows the 'ready' signal, you are to respond by saying the first word that comes to your mind other than the word that I have just spoken. Your response must be a single word, and you must say it just as quickly as you can."

If *S*, despite these instructions, repeats the stimulus word, he is cautioned not to do so, and the same stimulus is given again after several other stimuli have been used. If he continues to repeat the stimulus word during some 25 trials, *E* should forego further attempts to prevent this form of response. If *S* responds by a sentence or phrase, a compound word or a grammatical variation of the stimulus-word, he is similarly warned of this infringement of the instructions, and the stimulus words are similarly repeated later in the testing. In any event, the original response as well as the subsequent one had best be noted in the record, though the second ones should be used in computing the results.[3]

If any response seems incoherent, devoid of any apparent connection with the stimulus, ask *S* why he responded as he did, and make a note of his explanation.

As the test is somewhat wearisome with children, it is advisable to introduce a rest-pause of a minute or so after the 50th word, or even after the 25th, 50th and 75th words.

VARIATIONS OF METHOD.—(1) Use the stop-watch to measure the time elapsing between the stimulus and the response. Start the watch just as the stimulus is uttered; stop it when *S* utters his response; record the time in tenths of a second. This variation of method is strongly advised, despite the reasons advanced by Kent and Rosanoff for neglecting the measurement of the association time. The experience of other users of the Kent-Rosanoff test shows that the times are frequently valuable adjuncts in diagnosis. *E* must remember, of course, that the time

[3]With quite young children, say 4 or 5 years old, it will be impossible to follow these instructions precisely. Thus the Rosanoffs (18) were obliged in such cases to permit responses of a sentence form and to take the main word in the sentence as the desired single word.

does not always measure the speed of the association that is recorded, since between stimulus and response more than one mental process may intervene which is not reported by *S* and which may be quite unrevealed in the word he utters.

(2) Make the test without instructions for speed, but with explicit instructions to adopt a quiet, leisurely attitude in which the association is allowed to develop in whatever way it may. This method of conducting the association test yields responses that often differ widely from those obtained under instructions for speed, and it must be understood that the coefficients obtained from the frequency tables then possess no necessary correspondence with those obtained when the standard method is followed.

(3) Follow the suggestions just cited in Variation 2, with the additional proviso that *S* may respond by a phrase or compound word in case that be the first verbal association that rises in his mind. It is instructive to compare the responses obtained under this *Aufgabe* with those obtained by the standard method. Here, again, the calculated coefficients are not directly comparable with those established by Rosanoff with his tables and his instructions, though the method is, in the author's opinion, a better one for determining the degree of individuality in associative tendencies.

Treatment of Data.—To determine the coefficient of commonplaceness compare the responses for each one of the 100 terms with the responses listed in the Kent-Rosanoff frequency tables; record the several 'index-values,' then average them to obtain the coefficient. Thus, if to *table S* responds *chair,* the index is recorded as 267, because 267 of the 1000 persons tested by Kent and Rosanoff gave this response: if the association be *table-hard,* its index is 9; if it be *table-black,* the index is 0, because no one of the 1000 persons chanced to give that response. The association *table-black* and any other association which is not found in the frequency tables is termed an *individual* response, while any association found in the tables, whatever its index may be, is termed a *common* response. Any response that is a grammatical variant of a term listed in the tables is classed as a *doubtful* response, *e. g.,* the association

table-inky is doubtful since only *table-ink* is found in the list for that stimulus word.

The percentage of common, individual and doubtful responses are then computed. If no response is obtained (*e. g.*, "nothing," "don't know"), *E* may find it necessary to add a fourth class ("failures") to the three classes just mentioned.

The term *common response* is not entirely synonymous with the term *normal response*, because, obviously, the responses gathered from 1000 persons do not exhaust the possibilities of perfectly 'natural' associations. To meet this difficulty Kent and Rosanoff have given in their appendix statements that cover in a general way the responses that are to be deemed 'normal' for each of the 100 stimulus words and also still more general rules to cover associations to any stimulus word. These explanations (10, pp. 126-142) must be kept in mind whenever the question arises whether the responses of a given *S*, however individual they may be, are yet within the bounds of normality.

In certain of the words in their list this restriction of the indexes to those responses actually secured from the 1000 persons occasions a perfectly obvious and rather unfortunate artificiality. Take, for example, the word *city* (No. 79). As a response to it, 12 different cities have been named, and with the most divers frequencies, *e. g.*, *New York*, 99; Cleveland, 1. If *S* chanced to respond *Indianapolis*, he would have to be credited with an individual response, index 0. I would suggest that in this case all names of cities be counted together and the name of any city be given the resultant index, 124. Similar situations arise with other stimulus words, like *doctor*, *square*, *child*, *ocean*, etc., and with respect to the grammatical variants of many of the responses. Thus, for instance, the association *man-woman* has an index of 394, that *man-women* an index of 0. There are numerous such cases in which a very slight modification of the response alters enormously its index value, so that changes in the association which would appear psychologically indifferent remove responses from the realm of the commonplace and credit them with individuality.

Another criticism that might be raised against the use of the frequency tables in the manner prescribed is that the tendency toward commonplaceness may be abnormally raised by the chance giving of a very few responses whose index value is unusually high. To give the three responses *table-chair*, *dark-light*, *soft-hard*, alone, will give *S* 1059 points of commonplaceness, even if he should happen to give an entirely individual response to every one of the 97 remaining terms. To meet this difficulty, *S*'s may be compared simply with respect to the number of common and of individual responses they have given, or, as suggested by Bovet, by computing as an index of banality the number of terms to which the *most common* response is given and as an index of originality the number of unique responses. Still other methods have been suggested by Bovet.

In addition to computing the coefficient of commonplaceness, and the percentage of common, doubtful and individual responses and failures, *E* may attempt a further classification of the responses, and, indeed, he will find it quite desirable to do so whenever the number of individual responses distinctly exceeds the ordinary number, or whenever other features of the responses indicate the possibility of some anomaly in the associative processes. For this purpose the classification and analysis published by Kent and Rosanoff in conjunction with their frequency tables will serve satisfactorily.[4]

The following explanation may serve to assist in the use of this classification.

(a) While common responses are as a rule also normal responses, there are certain ones of them that may be termed *non-specific responses*, which, if present to an unusual degree, may have some pathological significance. A non-specific response is one which has so wide an application as to be a possible associate for almost any stimulus word, *e. g.*, such nouns as *thing, article, object*, or such adjectives as *good, small, useful, pleasant.*

Within the individual responses, in addition to individual non-specific responses (like those just mentioned, but not in the frequency tables), there may be distinguished:

(*b*) *Responses by sound*, that lead to neologisms, *i. e.*, construction of new words, *e. g., man-manion, anger-Angoria.*

(*c*) *Neologisms without sound relation*, as *dark-unbright, deep-deptableness.*

(*d*) *Repetition of the preceding response.*

(*e*) *Repetition of a response* five times or over (stereotypy), as the response *parent* to the stimuli *man, mountain, mutton, short, woman, cold*, etc.

(*f*) *Repetition of the preceding stimulus.*

(*g*) Response by derivatives, *i. e.*, grammatical variants of the stimulus word, *e. g., short-shortness, sweet-sweetened.*

(*h*) *Individual non-specific responses.*

(*i*) *Responses by sound*, but with actual words, *e. g., man-manners, short-shorthand.*

(*j*) Word complements, *i. e.*, responses in which an addition to the stimulus word forms a word, name or compound term in common use, *e. g., baby-hood, thirsty-blood, green-Paris.*

(*k*) *Responses by particles of speech*, as articles, numerals, pronouns, auxiliary verbs, adverbs of time, place and degree, conjunctions, prepositions and interjections, *e. g., chair-down, eating-sometimes, soldier-yours, whiskey-no.*

(*l*) *Association to preceding stimulus*, meaning a response not found

[4]The prolonged discussion concerning the most feasible and psychologically justifiable system of classifying responses in association tests is too lengthy for consideration here. For some account of recent classifications the reader may consult Wells (22) and Kelley (9). The idea of Kent and Rosanoff has been to forego logical classification in favor of a strictly empirical and objective system.

in the frequency tables for the word that it follows, but found in them as a response for the preceding stimulus, as in the pair of responses, *thief-night, lion-pocketbook.*

(*m*) *Association to preceding response*, meaning a response not found in the tables for the word that it follows, but found in them as a response to the response given for the preceding stimulus (whether in direct or reverse order), as in the pair of responses, *eating-table, mountain-floor.*

(*n*) *Repetition of a previous response* (distinguished from repetition of preceding response).

(*o*) *Repetition of a previous stimulus.*

(*p*) *Individual, but normal responses*, according to rules given in the appendix.

(*q*) *Association to a preceding response* (so judged by *E*, though neither response chances to be one of the 100 stimulus words), as in the pairs *priest-father, ocean-mother.*

(*r*) *Unclassified responses*—a rather large group in some types of pathological *S*'s, because of the presence of numerous incoherent responses, but also found with normal *S*'s when the response is affected by distracting circumstances, by purely personal experiences, etc.

In using this classification, responses that might be listed in two or more categories are to be assigned to the one of them cited earliest in the above list.

When times are obtained, the speed of each *S* is best indicated by the median, rather than by the average time of his 100 responses. Similarly, the best indication of variability is found in the quartile variation, *i. e.*, one-half of the difference between the 25th and the 75th time, when the times are arranged in order from fastest to slowest.

RESULTS.—(1) *Normal distribution* into the three fundamental categories of the Kent-Rosanoff system—common, doubtful, and individual—is best indicated by the results obtained by these workers for the 1000 *S*'s on which their frequency tables have been based. Their results are summarized in Table 58, wherein the distribution obtained by them and by other workers for other types of *S*'s has also been given to facilitate comparison. The point upon which most emphasis has been placed is the relatively small percentage of individual responses (6.8) given by normal *S*'s.

(2) *Normal times* for free association with the Kent-Rosanoff series have been reported by few experimenters. Miss Otis merely states that the times proved significant and valuable and that defective children were both slower and more variable than normal children. Goett (6), who used Jung's list, found that with normal children the mode was in the 2d second and

TABLE 58

Average Distributions for the Kent-Rosanoff Test (Compiled from Eastman, Kent, Rosanoff and Strong)

EXPERIMENTERS	S's	AGES	COMMON	DOUBTFUL	INDIVID'AL
Kent and Rosanoff	1000 normals	8–80	91.7	1.5	6.8
Kent and Rosanoff	247 insane	adults	70.7	2.5	26.8
Kent and Rosanoff	32 man.-dep.	adults	75.8	3.0	21.5
Strong	16 man.-dep.	adults	78.6	1.0	19.7
Eastman and Rosanoff	253 delinq't	11–17	84.0	2.2	13.2

was not much affected by age, while with mental defectives the mode was in the 3d second. Kelley's tests of 12 college students, with a list of 100 terms decidedly more difficult than the Kent-Rosanoff list, revealed a skewed curve with the mean slightly higher than the median, the median slightly higher than the mode; the mode was 1.0 sec., the minimal time 0.5 sec., the maximal time 3.5 sec. The author has found the average times for college students with the Kent-Rosanoff list to lie between 1.00 and 2.75 sec. An *average* less than 1.5 sec. may be construed as a fast association time.

(3) *Dependence on age.* That children give distinctly fewer common associations (and hence more individual associations) than adults is the general conclusion of all experimenters (Reinhold, Saling, Wreschner, Ziehen and the Rosanoffs), though Reinhold did not find the number of common responses to increase steadily from year to year, and the Rosanoffs believe that the differences between children and adults are practically obliterated after the age of 11.

Their results, expressed in per cents., are shown in Table 59. Graphs of these distributions will be found in the original text (p. 49). The increase in the frequency of individual responses at the age of 15 is attributed to the presence of a number of retarded pupils who were still members of a grammar school at this age. Failures to respond (sixth column), which include replies of "don't know," are found to be due usually to lack of familiarity with the stimulus words. While this conclusion is borne out by their detailed table of failures (18, p. 47), it

TABLE 59

Dependence of Distribution in the Kent-Rosanoff Test on Age
(Isabel Rosanoff and A. J. Rosanoff)

AGE	COMMON RESPONSES		DOUBTFUL RESPONSES	INDIVIDUAL RESPONSES	FAILURES TO RESPOND
	Specific	Non-Specific			
4	40.4	1.1	3.8	25.3	29.4
5	55.1	2.0	4.4	21.4	17.1
6	62.2	2.7	3.2	18.6	13.3
7	64.9	4.0	3.5	20.0	7.6
8	68.4	5.8	3.1	18.0	4.7
9	75.1	5.5	1.7	14.2	3.5
10	72.9	8.4	2.3	14.3	2.1
11	82.0	7.1	1.7	8.6	0.6
12	83.8	6.6	1.3	7.6	0.7
13	81.1	8.4	1.8	8.5	0.2
14	84.1	6.3	1.4	7.7	0.5
15	78.7	7.6	2.0	10.8	0.9
Adults	85.5	6.2	1.5	6.8	——

must be borne in mind that occasionally even normal adults reply "nothing," and again that young children often make no reply and yet subsequent questioning shows that they have had numerous visual and even verbal associates in consciousness to which for one reason or another they have given no expression.

A further analysis of these writers sheds some light upon the relative preponderance in children (300 cases, 4-15 years old) as compared with adults (86 normal cases, selected records containing not over 10 per cent. individual responses) of certain types of individual responses. Reference to this analysis, reproduced in Table 60, shows that the greater part of the excess individual responses given by children fall in the categories 'partial dissociation' and 'perseveration,' while the individual, but normal responses (by appendix to the frequency tables) are actually fewer with children.[5]

[5]In this analysis the term 'partial dissociation' embraces what have been described above as non-specific responses, responses by sound (including neologisms), word complements and particles of speech, while the term 'perseveration' embraces all varieties of responses to earlier stimuli or to earlier responses and repetitions of responses more than five times.

These experimenters argue that "it would seem, then, that the tendency of children to respond with individual reactions more

TABLE 60

Individual Responses of Children and Adults in the Kent-Rosanoff Test (Isabel Rosanoff and A. J. Rosanoff)

TYPES OF RESPONSE	86 NORMAL ADULTS	300 CHILDREN
Normal (by appendix)	41.8	20.0
Derivatives of stimulus words	0.3	0.1
Partial dissociation	8.0	11.1
Perseveration	6.1	27.8
Neologisms (without sound relation)	—	0.6
Unclassified	43.8	40.4

often than adults rests in a large measure upon a certain lack of mobility of attention which results in an inability to quickly dismiss from the mind previous stimulus or reaction words and to turn the mind wholly toward the new stimulus word."

The author is inclined to believe that here, as in not a few other instances in which children differ from adults in psychological tests, what we are really bringing to light is an inability of the children to understand the instructions or disinclination to follow them if they are understood. In other words, the regular instructions of the Kent-Rosanoff test constitute an artificial restriction of the natural associative tendencies, as will be shown further on. Adults are able and willing to maintain the proper attitude and follow the ruies of the game; many children are unable or unwilling to do so.

Again, as regards the perseverative tendencies, no instructions are given to the child to avoid repetition of association, while it is common for cultured adults to avoid repetition from some preconceived notion that they are called upon so to do.

Moreover, while data are lacking to prove this contention, it seems very likely that the amount of 'perseveration' witnessed in responses to an association test is much influenced by the speed with which the entire test is conducted, because the faster the succession of stimuli, the greater the 'hang-over' effects of the words, whether stimuli or responses, that have been in consciousness. Although precise statements are wanting, Rosanoff and his co-workers appear to have conducted their tests at a fast pace. It goes without saying that for purposes of comparison between any groups of *S*'s, the speed of giving the stimuli should be constant and that pauses for rest introduced with one group should be introduced likewise with the others.

A closely similar opinion is expressed by Kakise (8), when he says: "To sum up, these so-called characteristic forms in children and the abnormal can all be found in normal adults in their natural associations, *i. e.*, when they react according to natural and spontaneous suggestions,

as was the case with our experiment, and do not react according to artificial and 'sophisticated' associations, *i. e.*, by mere verbal associations, as is the case in the customary experiment with normal observers who are expert enough to obey the 'rules.'"

Another attempt to establish relations between age and associative type has been made by Miss Otis, who tested 200 normal children, aged 4 to 8 years, 40 in each of the 5 ages, and compared the results with those for 130 children in the Vineland, N. J., Training School for the Feeble-Minded, classified for mental age by the Binet-Simon tests.

TABLE 61

Types of Associative Response in Normal and Feeble-Minded Children (Otis)

Normals.

AGE	FAILURE	TYPE I	TYPE II	TYPE III	TYPE IV	TYPE V	TOTAL
4	1	21	4	0	11	3	40
5		6	13	1	14	9	43
6			2	6	13	18	39
7			2	1	8	29	40
8		1	1	3	5	30	40
All	1	28	22	11	51	89	202

Defectives.

AGE	FAILURE	TYPE I	TYPE II	TYPE III	TYPE IV	TYPE V	TOTAL
2		4					4
3		5	1			1	7
4		3	1		1		5
5		1			5	1	7
6	2	3	1		9	1	16
7			1		11	6	18
8				1	7	19	27
9		1			7	15	23
10					3	14	17
11						4	4
12						4	4
All	2	17	4	1	43	65	132

In Table 61, Type I signifies repetition of the stimulus, Type II non-logical responses (no apparent connection between stimulus and response), Type III responses by sound (whether by a real word or by a neologism), Type IV multiverbal responses (like *whistle—when you whistle, doctor—to make you better, table—there's a table*), and Type V normal responses (meaning here responses by one word, of which at least 50 per cent. must be found in the frequency tables). A child is classed as belonging clearly to one of these five types only when at least 50 per cent. of his responses are of the kind indicated, but the figures given in Table 61 include cases of 'mixed' types, which have been classed by the preponderant tendency. So far as normal children are concerned, it appears (1) that at 4 years more than half belong to the types characterized by repetition of the stimulus, (2) that non-logical responses (Type II) are characteristic of 5 years, (3) that multiverbal responses (Type IV) are very prevalent from 4 to 6 years, (4) that a normal type of response, in the sense here used, is established in 75 per cent. of children by the age of 8 years, though these children by no means respond like adults, or even like children of 12 when their detailed responses are taken into consideration.

The relation of speed of association to age is not so clearly made out as one might expect. Nearly all experimenters find that work with the Kent-Rosanoff lists takes longer with children than with adults. Ziehen concluded that free association times decreased markedly year by year and Wreschner reached a similar conclusion, but both Goett and Rusk report that there is no definite relation between speed and age, while Meumann calls attention to the fact that, though work progresses more rapidly with older children, the more intelligent not infrequently respond more slowly, and the less intelligent, by reason, seemingly, of their relatively less originality and paucity of imagery, frequently respond more rapidly. In the limited number of tests made by the author, children (of about the age of 9) have invariably been distinctly slower than adults.

(3) *Dependence on sex.* The results reported by Burt and by Burt and Moore show a number of inconsistencies: in one group at least (65 children in the Holt School, Liverpool) the

girls slightly exceeded the boys in number of associations written, while in another group (130 children, aged 13 years, in the Wallasey School) 35 per cent. of the boys exceeded the median of girls. Burt and Moore, in any event, conclude that "the males are far quicker than the females."

In their compilation of data from 1000 normal *S*'s Kent and Rosanoff did not find any considerable differences between the sexes in the nature of the distribution of the responses.

(4) *Dependence on practise.* Both Rusk and Wells (23) find that practise in giving free associations reduces the time. Since this practise is not gained by actual repetition of the same series of stimulus words, the gain in time must lie in facilitation of general factors that condition the process of associating. Wells finds that the responses become less emotional, that the number of supraordinate relations is diminished and that of simple language-motor responses is increased, while at the same time there is greater 'particularization' in the responses. Verbal connections appear to become 'loosened up,' and general linguistic readiness is augmented. The effect of practise, then, is to develop an easier, simpler and more superficial type of response. Practise also decreases the times, so that the median speed is reduced to about 1.2 sec. from any amount above that up to 3.0 sec., with the consequence that individual differences in speed are less after practise than before it.

(5) *Dependence on intelligence.* The original data collected by Kent and Rosanoff permit them to compare the responses of 100 persons of collegiate education with those of 100 persons of common school education. The comparison indicates (see their Table I, p. 9) more individuality in the responses of those of collegiate education, but the authors deem it unsafe to risk a definite generalization to this effect on account of the wide variability in individual records of both groups.

In the case of children Isabel and A. J. Rosanoff compared 21 'bright,' 21 'average' and 21 'dull' children (teachers' estimates) and secured the results shown in Table 62.

The same investigators contrasted 38 pupils who were pedagogically advanced with 38 pupils of the same ages who were pedagogically retarded, and found, similarly, that the retarded

TABLE 62

Distribution of Responses as Conditioned by Intelligence (Isabel and A. J. Rosanoff)

GROUP	COMMON	DOUBTFUL	INDIVIDUAL	FAILURES
Bright	79.0	3.4	12.0	5.6
Average	75.3	2.7	12.6	9.4
Dull	66.9	2.3	22.0	8.8

pupils gave more individual responses (13.7 vs. 9.8 per cent.) and the advanced pupils more common responses (86.9 vs. 81.6 per cent.). They believe that extreme departure from the distribution which is average for the age of the child in question is an indication of a fundamental difference in mental ability; that 'plus-variations' [exceptionally high percentage of common responses?] characterize cases of precocity, while "minus-variations border on the pathological."

These conclusions are distinctly at variance with those reached by Ziehen, by Wreschner and by Meumann, all of whom find a greater degree of originality, *i. e.*, more individuality, in the associations given by more intelligent children. Other differences cited by Meumann (14, 89-101) are the following: (1) the unintelligent more often misunderstand or misinterpret the stimulus word; (2) they more often fail to respond; (3) they give a greater number of incoherent and seemingly senseless associations; (4) they more often use very 'superficial' connections, such as grammatical variants of the stimulus, rimes, simple opposites; (5) they often give responses derived from phrases or verbal connections that they have learned in some school exercise; (6) they often exhibit an apparent precocity by giving responses like adults rather than the more concrete and pictorial associations that are characteristic of most children of their years; (7) they tend to stick to certain forms of response once they have begun to use them (perseveration). Reinhold, on the other hand, found that in two of four classes the better children showed more, and in the other two less originality than the poorer children: he also argues that no differentiation between intelligent and unintelligent children can be

made on the basis of the tendency toward responses by sound, as this tendency is found on repetition of the test to be quite variable and a mere matter of chance attitude or 'set.' Similarly, Winteler, who sought to distinguish two types of response (the one termed the perceptual or describing, the other the comparing or relating type), could discern no relation between intelligence and propensity to use these types.

(6) *Dependence on family relationship.* Fürst, who tested 100 persons in 24 families with Jung's test words (7) and classified the responses under various categories, concluded that persons related to one another tend to exhibit more similarity in the use of these types than do persons not related. He also concluded that the associative type of children resembles that of their mother more than that of their father.

(7) *The feeble-minded.* Miss Otis' results with Vineland children have been presented in Table 61, where it is shown that repetition of the stimulus (Type I) is a common tendency with low-grade mental defectives, that multiverbal responses (Type IV) are encountered more often and persist till a later age in feeble-minded than in normal children, and that normal responses (Type V) appear later and less regularly in the feeble-minded. Goett deems the test of diagnostic value for examining abnormal children. He states that imbeciles have slower association times, tend to repeat responses and to give an unusually large number of multiverbal and non-specific responses and responses of the 'predicative' type (*wood—burn, glass—breaking*). The 253 children examined by Eastman and Rosanoff seem to have been at least two years or more pedagogically retarded. The results accord quite closely with those of Miss Otis and of Goett in that they reveal an unusual number of non-specific responses, of repetitions of response, and of the use of particles. In addition, these investigators found a relatively large proportion of failures to respond, and a percentage of individual responses much above the average for normal persons, though not so great as in the insane (Table 58). On the other hand, really incoherent responses, senseless neologisms, etc., so frequent in the insane, were almost never given by the feeble-minded. These authors point out that the conclusions

just cited are true only for the group as a whole; a good many of the children rated as feeble-minded or delinquent gave normal associations. When to this admission is added the demonstration of Kakise, to which we have alluded, that multiverbal responses, repetition of the stimulus and other tendencies supposed to be characteristic of abnormal minds can also be found in normal adults, the value of the Kent-Rosanoff test as a device for diagnosis of individual cases is certainly much less evident than some of its friends would have us believe.

(8) *The insane.* That the insane show a relatively high frequency of individual responses is shown by the work of Kent and Rosanoff and of Strong (Table 58). Kent and Rosanoff have also shown by further analysis of their material (10, p. 29) that there are characteristic differences in the distribution of the various forms of individual responses in the several different forms of insanity, such as dementia praecox, paranoia, epilepsy, general paresis, manic-depressive insanity. A similar conclusion is reached by Ley and Menzerath (12). The results obtained by Strong with 16 cases of manic-depressive insanity show good agreement with those obtained by Kent and Rosanoff for 32 cases of the same sort. To what extent inferences may safely be drawn from peculiarities in the times of responses, particularly from excessive slowness of reply, with respect to the presence of hidded emotional complexes is a matter of much dispute.[6]

(9) *Dependence on instructions.* Attention has already been called to the difference in the outcome of the association test according as *S* is set to respond as *quickly* as possible or as *well* as possible.[7] It should be repeated that the conclusions drawn from the Kent-Rosanoff test with its frequency tables hold only when the instructions to respond by a single word as quickly as possible are strictly followed. In illustration reference may be made to the author's own responses, taken under Varia-

[6] A general idea of this problem may be gained from the references here cited from Jung, Ley and Menzerath and Levy-Sühl.

[7] Consult Meumann (13, 420 ff.) for further analysis of possible instructions for this test. Roels (17) has also called attention to the fact that *S*'s, despite uniformity of instructions, do adopt different attitudes toward the test and thus give different times and responses.

tion of Method No. 3 (leisurely response with permission to use phrases when such did appear first in consciousness). Three alterations are prominent when the responses are compared with those by the standard instructions. (*a*) The number of individual responses is decidedly increased, so that the coefficient of commonplaceness falls from 12.8 to 10.5 (reckoned in terms of the mean) or from 7.0 to 3.0 (reckoned in terms of the median). Striking examples are the following:

STIMULUS	QUICK RESPONSE	COEFFICIENT	LEISURELY RESPONSE	COEFFICIENT
soft	hard	365	pedal	0
needle	thread	160	stickpin	0
religion	faith	47	ecstasy	0
whiskey	rye	9	rotten	0
city	town	258	voleur[s]	0

(*b*) There are numerous responses by phrases, and, furthermore, comparison of these phrases with the single-term responses under standard instruction shows that the single terms were really picked out from the phrase that was rising in consciousness. That young children may not always stop to make this selection is the evident explanation of the tendency seen in them to respond by phrases, even despite repeated instructions to the contrary by *E*. The following examples will make this point clear:

STIMULUS	QUICK RESPONSE	LEISURELY RESPONSE
lion	beast	king of beasts
command	order	yours to command
justice	peace	justice, peace and mercy
child	father	child is father to the man

(*c*) There are numerous indications of 'perseverative' tendencies, especially in the use of the same response for a number

[s]By way of the phrase: "the city is full of thieves" and thence to the French for thief—an excellent illustration of the complexity of the process that may intervene before it is possible to ejaculate an oral response, and of the fact that the dropping of all introspective reports must rob the test of much of its interest.

of different stimuli, as *soft—pedal, smooth—soft, hard—soft, loud—soft pedal, quiet—soft pedal.*

REFERENCES

(1) P. Bovet, L'originalité et la banalité dans les expériences collectives d'association. *ArPs(f)*, 10:1910, 79-83.
(2) C. Burt, Experimental tests of higher mental processes and their relation to general intelligence. *JEPd*, 1:1911, 93-112.
(3) C. Burt and R. C. Moore, The mental differences between the sexes. *JEPd*, 1:1912, 273-284, 355-388.
(4) J. M. Cattell and Sophie Bryant, Mental association investigated by experiment. *Mind*, 14:1889, 230-250.
(5) F. C. Eastman and A. J. Rosanoff, Association in feeble-minded and delinquent children. *AmJIns*, 69: 1912, 125-141.
(6) T. Goett, Associationsversuche an Kindern. *Z. f. Kinderheilkunde:* 1911. (Reprint, Pp. 105.) For detailed review, see *ArPs(f)*, 12: 1912, 192-193.
(7) C. G. Jung, The association method. *AmJPs*, 21: 1910, 219-269.
(8) H. Kakise, A preliminary study of the conscious concomitants of understanding. *AmJPs*, 22: 1911, 14-64.
(9) T. L. Kelley, The association experiment: individual differences and correlations. *PsR*, 20: 1913, 479-504.
(10) Grace H. Kent and A. J. Rosanoff, A study of association in insanity. *AmJIns*, 67: 1910. (Reprint. Pp. 142.)
(11) M. Levy-Sühl, Ueber experimentelle Beeinflussung des Vorstellungsverlaufs bei Geisteskranken. Leipzig, 1911. Pp. 142.
(12) Ley et Menzerath, Les associations des idées dans les maladies mentales. Gand, 1911. Pp. 200.
(13) E. Meumann, Vorlesungen zur Einführung in die exp. Pädagogik. Vol. II. Leipzig, 1913. (2d ed.)
(14) E. Meumann, Intelligenzprüfungen an Kindern der Volksschule. *EPd*, 1: 1905, 35-101.
(15) Margaret Otis, A study of association in defectives. To appear in *JEdPs* (?).
(16) F. Reinhold, Beiträge zur Associationslehre auf Grund von Massenversuchen. *ZPs*, 54: 1910, 183-214.
(17) F. Roels, La recherche du mot de réaction dans les expériences d'association. Extrait, *Annales de l'Institut Superieur de Philosophie* (Louvain), 3: 1914, 553-573.
(18) Isabel Rosanoff and A. J. Rosanoff, A study of association in children. *PsR*, 20: 1913, 43-89.
(19) R. R. Rusk, Experiments on mental association in children. *BrJPs*, 3: 1910, 349-385.
(20) Gertrud Saling, Associative Massenversuche. *ZPs*, 49: 1908, 238-253.
(21) E. K. Strong, Jr., A comparison between experimental data and clinical results in manic-depressive insanity. *AmJPs*, 24: 1913, 66-98.
(22) F. L. Wells, (*a*) A preliminary note on the categories of association reactions. *PsR*, 18: 1911, 229-233. (*b*) The question of association types. *PsR*, 19: 1912, 253-270.
(23) F. L. Wells, Practise effects in free association. *AmJPs*, 22: 1911, 1-13.
(24) J. Winteler, Experimentelle Beiträge zu einer Begabungslehre. *EPd*, 2: 1906, 1-48, 147-247.

(25) R. S. Woodworth and F. L. Wells, Association tests. *PsMon*, 13 : 1911 (No. 57). Pp. 85.
(26) A. Wreschner, Die Reproduktion und Assoziation von Vorstellungen. Leipzig. *Ergänzungsband, ZPs*. 3 : 1907.
(27) T. Ziehen, Die Ideenassoziation des Kindes. Berlin, 1898 and 1900.

TEST 34

Controlled association: logical relations.—These tests differ from the preceding test of association in that they demand the giving of a response which is so restricted that only a very limited number of terms may be deemed correct associates. There are, of course, numerous forms of controlled association, since numerous logical relations may be demanded between the stimulus words and the responses. The relations that have received most attention in the literature of mental tests are part-whole, genus-species (subordinate) and opposites. Other less often used relations are whole-part, agent-action (subject-verb), action-agent (verb-subject), attribute-substance (adjective-noun), substance-attribute (noun-adjective), cause-effect, effect-cause, species-genus (supraordinate), co-ordinate and mixed relations.

Just precisely what mental capacities are measured by these tests is not always clear. Of course, it may be said roughly that they call forth the "ability to appreciate relationships and to control associations." It is also evident that the skill in handling these various relations is based upon what is known in psychology as a "determining tendency," or "adjustment to react according to instructions," and that "the more completely this adjustment dominates the performance, facilitating the right responses and inhibiting other, interfering associations and perseverations, the less hesitation and confusion will occur and the more prompt will be the reaction."

On the other hand, an obstacle both to designating the capacities measured and to evaluating the results of these tests lies in the selection of the stimulus words themselves, for, if the terms are too difficult, failures appear due to lack of familiarity with their meaning or with the meaning of the terms connected with them in various logical relations; while, if they are too simple, no 'thinking' is demanded and the responses are

given well-nigh automatically. Moreover, the inclusion in a list of terms of one or more stimuli that are markedly different from the others in this respect introduces a source of difficulty in administering the test that is hard to meet, especially in group tests. The only solution of these difficulties is to discover by comprehensive testing what might be termed the 'association value' of each stimulus word for *S*'s of a given sex, age, intelligence, etc., and then to prepare standardized lists of stimuli suited by their like association values to the measure of controlled association in specified types of *S*'s. Much has been accomplished in this direction, but much still remains to be done.

Consideration of the various possible forms of controlled association is limited in what follows mainly to the most-used relations, part-whole, genus-species, and opposites.[1]

A. THE PART-WHOLE TEST

MATERIALS.—(1) For individual tests: Split-second stop-watch. Set of 20 cards (and 3 samples), each containing a stimulus word. Paper for recording times, responses and remarks. (2) For group test: Stop-watch or special seconds clock. Printed form containing the same stimuli and provided with spaces for the recording of the 20 associates.

The terms incorporated in these cards and in the form are those recommended by Woodworth and Wells as the result of numerous efforts at standardization. Cards are used, however, instead of the narrow cardboard strip of these authors in order that the time of each response may be measured by itself. The paper form is used to admit of written group tests.

The terms proposed by Pyle for this test are: window, leaf, pillow, button, nose, smokestack, cogwheel, cover, letter, petal, page, cob, axle, lever, blade, sail, coach, cylinder, beak, stamen. His supplementary list is the same as that of Woodworth and Wells.

The 10-word lists used by Rusk were: ear, wheel, beak, inch, platform, mast, branch, kernel, funnel, buckle: *alternatives*, mouth, handle, claw, ounce, pavement, sail, stem, core, boiler, knob.

The ten terms employed by Miss Norsworthy were: door, pillow, letter, leaf, button, nose, cover, page, engine, glass.

The ten terms employed by Wyatt are not specified by him.

---

[1]The mixed relations or analogies test is dealt with separately as Test 34A. For further details concerning other tests of logical relationship, consult Rusk (15), Watt (20) and Woodworth and Wells (22).

METHOD.—Instruct *S* as follows: "Each one of these cards has printed on it a word. As soon as I uncover a card I want you to look at the word on it and then, as quickly as you can, say aloud the name of the whole thing of which that word is a part. The word you read is a part: you are to name the whole. For example, if the card should have the word *fur* on it, you might say *cat* or *seal* or *fox*. We will try these sample cards first to make sure you understand."

After a warning 'now' remove the cover-card from the top of the pile and take *S*'s time for the first sample—*button*. Follow with the other samples, *leaf* and *drawer*. These cards are displayed and the time taken just as in the test proper, in order to accustom *S* to the regular procedure. Misunderstandings are, of course, corrected and cleared away.

Proceed with the 20 standard test cards. Record on the blank sheet of paper the times, in tenths of a second, together with *S*'s responses and any comments that suggest themselves. Each card is provided with its own cover-card. They are best removed with the left hand and the watch started simultaneously with the right. It is recommended that the split-second watch be used, so that one hand may be stopped when *S* first responds; then, if his response chances to be wrong, say 'No, give me another,' and take the time of his second attempt with the other hand of the watch. In this event, both times and both responses are recorded.

VARIATIONS OF METHOD.—(1) Individual testing may also be carried out, especially if none of the terms is likely to cause unusual delay, by giving *S* the printed form provided for group tests and taking his total time for naming *orally* the entire series of responses.

(2) For a group test by the work-limit method[2] (which is recommended for *S*'s who are competent to record their own time) use the printed forms and the special seconds clock, after the samples above mentioned have been displayed on a blackboard and discussed with the *S*'s. The clock is started at the signal for turning over the forms. Each *S*, of course, makes his own written record.

[2]See Vol. I, p. 8, Section (7).

(3) For a group test by the time-limit method use the same forms and stop all *S*'s at a time-limit so chosen on the basis of preliminary trials with *S*'s of that grade of ability that the fastest *S* shall reach about the 16th term on the list.[3] It is hardly necessary to state that the scores obtained by either group test are not directly comparable with those obtained by the individual method.

TREATMENT OF DATA.—(1) In the individual test the best indication of speed is furnished by the median. For a measure of variability the semi-quartile variation may be used, *i. e.*, one-half the difference between the 5th and 15th time, when the series of times is arranged in order from fastest to slowest. If the *S*'s are competent, the errors will ordinarily be negligible, so that performance may be measured in terms of speed alone. If it should happen that differences in speed are slight, while qualitative differences are well-marked, speed may be neglected and performance rated in terms of quality, as by scoring 1 for each well-chosen associate, 0.5 for each 'partly right' associate, and 0 for wrong associates or omissions. If both speed and correctness need to be considered, some of the methods suggested in the cancellation test (No. 26) or in the opposites test (below) may be employed.

(2) When individuals are tested by recording the total time for the entire list (Variation 1), errors may again be neglected if few and of slight moment; if more serious, the time may be increased by adding to it a penalty figured on the basis of the average time taken to utter a correct response to each stimulus omitted or responded to wrongly.

(3) In group tests by the work-limit method, performance may, similarly, be taken in terms of total time, or of correctness, or of some combination of time and correctness.

(4) In group tests by the time-limit method, the simplest method of scoring is that of crediting 1, 0.5 or 0 for each re-

[3]The lists of terms in these tests of controlled association have been so arranged by Woodworth and Wells that the terms lying between the 8th and the 16th represent as nearly as possible stimuli of equal difficulty. Thirty sec. will suffice for testing competent adults. Pyle recommends 60 sec. for Grades 2, 3 and 4; 45 sec. for other grades (and 30 sec. for adults?).

sponse, as above explained. If necessary to compare the performances of groups that have had different time-limits, they may be related by computing them all as if 60 sec. had been assigned, *e. g.*, by multiplying the score of adults by 2, etc.

RESULTS.—(1) Norms of performance in the part-whole test are supplied chiefly in the data published by Woodworth and Wells, by Miss Norsworthy and by Pyle. Tests of adult college students by the first-named authors show that the average association time for this variety of controlled association (Variation of Method No. 1) may be taken as 1.53 sec., P.E. .06, with a range for different individuals of from 1.03 to 2.50 sec. The median times reported by Rusk for 22 children, aged 7 years 6 months to 14 years 9 months, under Meumann's 'B' instructions (emphasizing quality rather than speed) range from 1.6 to 5.0 sec. Miss Norsworthy's norms, based on 504 cases, represent results with her list of 10 words, no time-limit, scored in terms of number of correct associates. Pyle's norms are based upon his list of 20 words, scored in number correctly written in a group test, computed on a basis of 60 sec. time-limit.

TABLE 63

Performance in the Part-Whole Test (Norsworthy)

AGE	8	9	10	11	12	13	14	15	16	ADULTS
Median	6.5	7.8	7.8	8.7	8.7	9.0	9.0	9.0	9.0	10.0
P. E.	2.3	1.3	1.9	1.1	1.2	0.7	0.7	0.7	0.7	0.5

TABLE 64

Correct Associates Written in 60 Sec. Part-Whole Test (Pyle)

SEX	AGE	8	9	10	11	12	13	14	15	16	17	18	ADULT
Male	Cases	31	67	70	65	76	77	62	42	35	12	23	66
	Aver.	5.5	6.5	7.3	8.9	8.9	11.1	12.2	14.8	15.9	15.8	19.3	18.5
	A. D.	3.6	2.9	2.5	2.8	3.4	4.3	4.1	5.5	5.3	4.0	5.6	3.6
Female	Cases	43	64	88	67	87	71	63	48	51	38	28	87
	Aver.	4.6	5.9	7.8	10.5	10.0	10.8	12.5	14.0	16.9	16.2	19.7	19.7
	A. D.	2.6	2.4	2.9	3.5	3.7	3.5	3.2	4.5	4.5	4.8	4.6	3.4

(2) *Dependence on age.* The results obtained by both Pyle and Miss Norsworthy show that performance in this test undergoes a fairly steady improvement from 8 to 18 years. The lack of any correspondence between speed and age reported by Rusk is probably due to the small number of cases examined by him.

(3) *Dependence on sex.* Sex-differences are not sufficiently evident to warrant conclusions, though it may be surmised that girls and women tend to be slightly superior to boys and men.

(4) *Dependence on intelligence.* Wyatt found a fair degree of correlation with intelligence (0.67, P.E. .07 in one group using teachers' estimates, and 0.56, P.E. .08 in another group, using class examinations as the basis for intelligence).

(5) *Feeble-minded.* The work of Miss Norsworthy shows that mentally defective children are distinctly inferior to normal children in this test: thus the percentage of normal children with a record above the median, above —1 P.E., and above —2 P.E., would, of course, be 50, 75, and 91, respectively, but the percentages of feeble-minded children obtaining these three grades of efficiency were but 9, 17, and 27, respectively. That is, only 9 per cent. of the feeble-minded children reached the degree of efficiency attained by one-half of the normal children, etc.

(6) *Other correlations.* Wyatt obtained with his Group I a moderately satisfactory coefficient of reliability, 0.65. His correlations with other tests range from 0.09 to 0.77. The lowest correlation was with the letter-squares test; the higher correlations appeared with analogies (0.67), the completion test (0.75) and word-building (0.77).

B. THE GENUS-SPECIES TEST

MATERIALS.—(1) For individual tests: Split-second stop-watch. Set of 20 cards (and three samples) each containing a stimulus word. Paper for recording times, responses and remarks. (2) For group tests: Stop-watch or special seconds clock. Printed form containing the same stimuli and provided with space for recording the 20 associates.

These 20 terms are those recommended by Woodworth and Wells.

The 10 terms used by Miss Norsworthy are: book, tree, room, toy, name, dish, boat, game, plant, fish.

The 20 terms proposed by Pyle are: mountain, city, weed, metal, furniture, machine, author, planet, river, book, ocean, fruit, country, animal, bird, food, lake, tool, fish, money. His supplementary list is the same as Woodworth and Wells.

The 10-word lists used by Rusk were: tree, fish, college, battle, picture, tool, hero, lesson, taste, wrong: *alternatives*, bird, leaf, game, poem, song, toy, hobby, book, smell, virtue.

METHOD.—Instruct *S* as follows: "Each one of these cards has printed on it a word. As soon as I uncover a card, look at the word on it and then, as quickly as you can, say aloud the name of some particular thing that belongs in the class that is given on the card. The word you read is the name of a class or genus; you are to name an example of that class, a species of that genus. For example, if the card should have on it the word *taste,* you might say *sweet* or *salt,* or if the word *verb,* you might name any verb like *run* or *go.* We will try three sample cards first to make sure you understand." Follow the procedure outlined for the part-whole test in regard to the use of the sample cards, timing, etc.

VARIATIONS OF METHOD.—Follow the suggestions given for part-whole test, save that here the samples will be *bird, dish* and *game.*

TREATMENT OF DATA.—Follow the suggestions given for the part-whole test.

RESULTS.—(1) Tests of adult college students by Woodworth and Wells (Variant Method No. 1) show for the genus-species test an average association time of 1.84 sec., P.E. .07, with a range for different individuals of from 1.20 to 2.63 sec. The medians reported by Rusk for 22 children from about 7 to 15 years of age, with quality emphasized more than speed, range from 1.6 to 11.4 sec. The norms reproduced here from Miss Norsworthy are based on 511 cases and represent performances made with her list of 10 words, no time-limit, scored in terms of number of correct associates. The norms reproduced from Pyle are based on his list of 20 words, scored in terms of number correctly written in a group test, computed on a basis of 60 sec. time-limit.

TABLE 65

Performance in the Genus-Species Test (Norsworthy)

AGE	8	9	10	11	12	13	14	15	16	ADULTS
Median	5.0	5.0	7.0	9.2	9.2	9.3	9.3	9.5	9.5	10.0
P. E.	2.0	2.7	2.9	1.9	0.7	0.4	0.5	0.5	0.5	0.0

TABLE 66

Correct Associates Written in 60 Sec. Genus-Species Test (Pyle)

SEX	AGE	8	9	10	11	12	13	14	15	16	17	18	ADULT
Male	Cases	29	67	66	62	69	68	64	41	33	18	16	65
	Aver.	4.6	5.7	6.5	7.2	7.1	10.0	10.5	11.1	15.2	14.0	17.3	15.1
	A. D.	3.4	3.4	3.7	3.3	2.5	3.8	3.8	5.4	4.3	4.1	6.0	4.0
Female	Cases	34	65	84	63	81	64	55	40	45	32	25	86
	Aver.	5.5	5.4	7.8	8.2	9.3	9.5	11.8	14.0	16.4	16.0	18.3	15.5
	A. D.	3.6	2.5	3.2	3.7	2.9	3.2	3.2	4.2	5.4	4.9	5.3	3.8

(2) *Dependence on age.* Though Rusk can find no definite relation between speed and age in the genus-species test, the figures reported by both Pyle and Miss Norsworthy show a general improvement with age, despite certain exceptions. Miss Norsworthy's test was obviously too easy for ages of 11 and above, so that any tendency toward improvement beyond 11 was obscured. Rusk's negative result is explicable partly by his instructions against haste and partly by the small number of cases he tested.

(3) *Dependence on sex.* Pyle's averages make it reasonable to assume a slight superiority of girls over boys, since they show this superiority in ten of the age groups.

(4) *Feeble-minded children,* according to Miss Norsworthy's results, are distinctly inferior in this test to normal children of the same age: only 9 per cent. reach the median of the normal children; only 16 per cent. reach —1 P.E.; only 17 per cent. reach —2 P.E. of normal children of their age.

C. THE OPPOSITES TEST

MATERIALS.—(1) For individual tests: Split-second stop-watch. Two sets of 20 cards each (exclusive of samples), one set of easy, and one of moderately difficult stimulus words. Paper for recording times, responses and remarks. (2) For group tests: Stop-watch or special seconds clock. Printed forms containing the same stimuli (one for easy and one for difficult words) and provided with spaces for recording the 20 associates.

The opposites test has been extensively used and has appeared in a variety of forms. The most common lists are printed herewith. They demand a few words of explanation.

Lists I, II and III represent the standardized lists for easy opposites prepared by Woodworth and Wells: Lists I and II, which are those used by Briggs (his Tests 43 and 44) are presumed to be of equal difficulty and to be so arranged that the last half is just as difficult as the first half; List III. which is the set of easy opposites here recommended, is a selection of the 20 easiest opposites in Lists I and II.

MATERIAL USED BY VARIOUS INVESTIGATORS IN THE OPPOSITES TEST.

I	II	III	IV	V
long	north	high	good	best
soft	sour	summer	outside	weary
white	out	out	quick	cloudy
far	weak	white	tall	patient
up	good	slow	big	careful
smooth	after	yes	loud	stale
early	above	above	white	tender
dead	sick	north	light	ignorant
hot	slow	top	happy	doubtful
asleep	large	wet	false	serious
lost	rich	good	like	reckless
wet	dark	rich	rich	join
high	front	up	sick	advance
dirty	love	front	glad	honest
east	tall	long	thin	gay
day	open	hot	empty	forget
yes	summer	east	war	calm
wrong	new	day	many	rare
empty	come	big	above	dim
top	male	love	friend	difficult

VI	VII	VIII	IX
day	great	succeed	tender
asleep	hot	strict	animated
absent	dirty	tardy	proficient
brother	heavy	sleepy	impoverish
best	late	suspicious	cruel
above	first	rigid	generous
big	left	suave	haughty
backwards	morning	sinful	silly
buy	much	conservative	insignificant
come	near	refined	disastrous
cheap	north	pride	miser
broad	open	despondent	result
dead	round	imaginary	hindrance
land	sharp	beautiful	strength
country	east	injurious	innocent
tall	known	diligent	busy
son	something	sell	remember
here	stay	sure	increase
less	push	active	preserve
mine	nowhere	venturesome	belief

List IV, one of the oldest and most employed, appears in several published articles and texts by Thorndike; it forms one of Simpson's easy opposites, has been used by Miss Norsworthy in an extensive study, and constitutes the regular test list prescribed by Pyle. The opposite of this list, *i. e.*, bad, inside, etc., has been used by Bonser, by Miss Norsworthy and by Mrs. Squire.

List V is proposed by Pyle as harder opposites for use with adults.

List VI has been used by Bonser, by Mrs. Squire and (with two changes) by Simpson.

List VII has been used by Bonser, by Mrs. Squire and (with some changes) by Simpson and by Carpenter. It also appears in Thorndike's tests.

Lists VIII and IX are two of four hard opposites used by Simpson.

The easy opposites test may be regarded as fairly well standardized so far as choice of material is concerned. But List III, which has been selected as best for younger children, will prove too easy for most *S*'s of 10 years or over, and we have no lists of moderately difficult and very difficult opposites that have been tested by very extensive experimentation. The set proposed by the author has been selected from the 50 terms used by Hollingworth, who, in turn, selected them from a list of 200 tested by Woodworth and Wells and showing association times of from 2 to 5 sec. The attempt has been made on the basis of the author's trials with college students and with the assistance of Dr. Hollingworth to select 20 hard opposites that shall be relatively easy to score and that shall be of closely similar difficulty.

METHOD.—Use the easy opposites for younger children, the more difficult ones for children over 10 or thereabouts and for adults. Instruct *S* as follows: "Each one of these cards has printed on it a word. As soon as I uncover a card, look at the word on it and then, quickly as you can, say aloud a word that

means just the opposite to it. For instance, if the card should have on it the word *dirty,* you would say *clean.* We will try three sample cards first to make sure you understand." With the hard opposites follow the procedure outlined for the part-whole test with regard to the use of the sample cards, timing, recording responses, etc.

Variations of Method.—Follow the suggestions given for variations of method in the part-whole test, with due regard for changes in sample terms, etc.

When the harder opposites are used, particularly with adults, the instructions may be altered to put special stress upon the giving of an exact opposite, *i. e.,* the emphasis may be placed upon quality rather than upon speed of performance. Further, it is well to instruct *S* that opposites formed by the use of the prefixes *un* or *in,* or of the suffix *less,* will not be allowed, save when the root of the stimulus word is changed; thus, for instance, *inharmonious* would not be accepted for *harmonious,* nor *unsafe* for *safe,* but *harmless* would be accepted for *dangerous.* The split-second watch may then be used to advantage by rejecting responses that are incorrect, and recording the time of various responses made before the proper one is given.

To test the effect of practise, the cards may be shuffled and the series repeated any desired number of times, as in the procedure adopted by Hollingworth in his tests of the effect of caffein.

Treatment of Data.—This may follow the directions given for the part-whole test. With older children and adults, however, when the instructions have emphasized quality rather than speed, the scoring of responses should be decidedly rigorous, and it may be well for many purposes to permit only a single correct opposite and allow nothing for 'partly correct' responses.[4]

---

[4]Mrs. Squire counted as errors all responses that were not accurate and gave no credit for adverbs when adjectives were correct, nor even for approximate opposites, urging that "there is no mental test in which an approximate is less permissible."

The important thing is, of course, that *E* should settle upon the type of mental activity that is to be demanded (speedy approximate responses or rigorous precision) and then adjust instructions and scoring to measure this aspect of the associative process. In using the cards

RESULTS.—(1) *Norms.* The average time of response of adult college students is given by Woodworth and Wells as 1.11 sec., P.E. .04, range 0.85 to 1.40, for the easy list. The norms obtained by Pyle, Miss Norsworthy, Mrs. Squire and Carpenter are reproduced in Tables 67, 68, 69 and 70, respectively: these figures may not be compared directly with one another on account of differences in materials, scoring and other conditions, but they will serve as bases for conclusions with regard to the dependence of performance in the test upon age, sex and other factors.

Pyle's norms represent the average number of opposites that could be written in 60 sec. in a group test, using a list formed of the opposites of List IV, above.

Miss Norsworthy's figures refer to the number of correct associates to the opposites of List IV (her First List) and to List IV itself (her Second List), given by about 611 normal children of both sexes. Here the maximal possible score is evidently 20.

Mrs. Squire's figures show the average time in sec. and the average number of correct responses (rigid scoring) for small groups of pedagogically unretarded children for three separate lists (our Lists VI, VII and the opposites of IV, respectively). For the standards proposed by her for each age, see the original article, pp. 500-506.

Carpenter's results are based upon what he describes as practically the same terms as our List VII. The figures show the errors and the

TABLE 67

Correct Associates Written in 60 Sec. Opposites Test (Pyle)

SEX	AGE	8	9	10	11	12	13	14	15	16	17	18	ADULT
Male	Cases	33	65	60	61	72	65	61	40	33	17	22	62
	Aver.	9.0	8.4	7.5	10.9	11.5	14.5	14.5	16.0	18.6	17.6	22.4	22.1
	A. D.	3.3	3.0	3.1	2.9	2.9	4.5	4.3	5.2	5.3	3.3	3.2	3.3
Female	Cases	33	56	77	65	74	73	58	49	48	27	26	85
	Aver.	8.0	7.6	10.9	11.2	13.9	14.9	17.4	17.3	19.3	21.4	23.4	23.4
	A. D.	4.0	2.9	3.1	3.6	3.6	4.3	3.9	5.1	4.2	4.9	3.1	4.0

for individual testing it will be found convenient to list upon the back of each card the different words that are given by *S*'s, together with the score previously determined upon for each word.

For a more elaborate system of equating speed and quality of work the reader may consult Simpson (16, pp. 14, 16).

The work of Woolley and Fischer contains many valuable suggestions for the evaluation of various responses in the opposites test (see 22a, pp. 216-221).

time in sec. obtained in individual tests in which the terms were read by *E* and responded to orally by *S*. His records are confessedly "of little value below age 9." Since the times were taken with an ordinary watch and merely express the total time occupied by the test for each child, it is impossible, of course, to derive from them any precise idea of the association time, strictly speaking.

TABLE 68

Correct Associates of a Possible 20. Opposites Test (Norsworthy)

LIST	AGE	8	9	10	11	12	13	14	15	16	ADULTS
Opp. of IV.-- -----	Median	7.4	9.0	9.9	12.5	13.5	14.0	14.5	15.0	15.5	20.0
	P. E.	2.0	2.0	3.0	3.0	2.6	2.5	2.3	2.3	2.3	1.0
IV.----------	Median	8.7	9.5	11.5	13.1	14.7	16.4	17.8	18.5	19.0	20.0
	P. E.	1.4	1.7	2.2	2.9	3.6	2.4	2.0	2.0	2.0	1.0

TABLE 69

Correct Associates and Times. Opposites Test (Squire)

AGE	LIST VI		LIST VII		LIST IV (OPP.)		AVER. FOR ALL	
	Words	Time	Words	Time	Words	Time	Words	Time
6	8.5	192.3	11.3	143.3	11.6	120.0	10.5	151.8
7	13.1	155.6	15.0	137.7	14.5	117.6	14.2	136.9
8	16.1	110.3	17.1	98.2	16.7	104.9	16.6	104.8
9	17.6	103.7	16.5	101.1	17.7	98.4	17.2	101.7
10	17.1	87.1	17.7	87.0	18.2	76.2	17.6	83.4
11	19.3	79.3	19.0	102.0	19.6	68.3	19.3	83.2
12	19.5	81.2	19.2	85.5	19.2	63.2	19.3	63.2
13	19.4	72.5	19.2	61.5	19.2	65.0	19.3	66.3

TABLE 70

Errors and Times. Opposites Test (Carpenter)

AGE	7	8	9	10	11	12	13	14
Cases ----------	7	19	46	50	41	44	58	49
Aver. Time-----	118	118	108	101	98	82	79	71
Aver. Errors ---	4.1	4.2	4.5	4.7	4.8	3.8	3.6	3.2

(2) *Dependence on age.* The evidence from these four tables, as well as from the results of Woolley and Fischer, though not guaranteeing in every instance a uniform increase in quantity or quality of performance from year to year, undoubtedly justifies the generalization that both speed of finding opposites and accuracy of the responses made, improve with age. The failure of this tendency to appear in the tables published by Bonser is apparently due to the inadequacy of his test material for bringing out the abilities of the older pupils.

(3) *Dependence on sex.* In all comparisons of groups the superiority of females over males is readily noted. Thus, girls surpass boys in 10 of Pyle's 12 age-groups; Hollingworth (9) found women faster than men in naming opposites both before and after practise; Burt and Moore report that in one group only 29.2 and in another only 42.2 per cent. of the boys reached the median mark of girls; Bonser found girls superior to boys in every school grade, though it is significant that in most of his groups more boys than girls were found in the highest quartile. Woolley and Fischer report a slight superiority of girls, but add that "it is too small to be considered very significant."

(4) *Dependence on school grade.* Bonser found a general progress from grade to grade, coupled with a decrease of variability; his results are somewhat affected by the fact that his test was too easy for use much beyond the 6A grade. On the other hand, the curves of distribution published by Chambers show that it is impossible to distinguish 7th grade and 8th grade pupils by their curves of distribution in this test.

(5) *Dependence on intelligence.* With the exception of Winteler's conclusion (based on the study of only 8 boys, divided into two contrasted groups) that the opposites test is less well adapted than other forms of controlled association to reveal differences in intelligence, the general opinion of experimenters is decidedly favorable to its use for this purpose. Mrs. Squire found that pedagogically retarded pupils show a distinctly lower general average performance with more irregularity in speed and quality of work than the unretarded. Bonser ob-

tained a correlation of 0.85 between opposites and the average standing in all the tests used by him to measure ability to reason, and declares that "the opposites test seems to be a test of rather superior merit as a single test for this general form of mental ability." In Simpson's interesting study of two contrasted groups of adults (17 of superior, and 20 of quite inferior general ability) the two groups were completely separated by both the easy and the hard opposites test, *i. e.,* no person in the poorer group did as well as the poorest in the good group. Simpson estimates the true correlation with the intelligence of people in general to be as high as 0.82 for the easy, and 0.96 for the hard opposites test.

At Bedford Hills Reformatory, Miss Weidensall reports a correlation of 0.79, P.E. .03, between rank-order in opposites and the estimate of intelligence of the women made by the director of the Industrial School of the institution. The correlation would have been higher under more favorable conditions for the testing and the estimating. Again, the institutional women who had received the most schooling showed almost exactly the same ability as the Cincinnati 15-year-old working girls tested by Woolley and Fischer, whereas the Bedford Below-Grade Group (schooling less than Grade 5B) were decidedly inferior to the Bedford Grade Group and to the Cincinnati girls.

(6) *The feeble-minded and delinquent.* No feeble-minded child, according to Miss Norsworthy's figures, reached the median performance of normal children; only about one in a hundred were better than —1 P.E. and only about six in a hundred were better than —2 P.E. of normal children of their age.

Reference has just been made to the comparison of delinquent women and school girls. Dr. Weidensall's results show also that the Bedford women, taken as a group, are slightly inferior to the Cincinnati 15-year-old working girls at the upper quartile, 7.5 per cent. less accurate at the median and 26.2 per cent. less accurate at the lower quartile. It is of interest to note that the opposites test proved somewhat difficult of comprehension for these *S*'s, so that special explanations had to be contrived and repeated as well before the testing could proceed.

(7) *Dependence on practise.* Hollingworth (9) put 11 men and 8 women through 100 trials with opposites, and also tested their speed in reading the stimuli and the responses from a typewritten sheet. The results were:

Initial naming (average of trials 2-6),	men	113.5	sec.,	women	99.7	sec.
Final naming (average of trials 96-100),	"	36.3	"	"	31.2	"
Reading directly (average of 5 trials),	"	18.5	"	"	16.1	"

It is seen that extensive practise increases the speed of the associative process markedly, but that even after 95 trials a considerable part of the time taken in the test is occupied by the process of association as over against the time needed in reading and uttering the words. The speed attained by individuals in the test by reading correlates with the speed obtained in the regular test of naming the opposites by approximately 0.60.

(8) *Dependence on fatigue.* In the course of his experiments upon the effects of caffein Hollingworth (8a) was able to observe the effects of time of day upon the opposites test under unusually favorable conditions as regards elimination of the practise error. In preliminary experiments (tests made at intervals of two or three hours) and also in more intensive experiments (15 trials between 10.30 A. M. and 10.30 P. M.) there appeared a distinct reduction in speed of naming opposites as the day passed, and the fatigue effect was more pronounced in this test than in any others that were tried. After the second trial the initial records of the day were never surpassed.

(9) *Dependence on race.* In opposites, and also in other controlled association tests (genus-species, part-whole), Pyle (13a) found negro children of both sexes less than half as efficient as white children.

(10) *Reliability.* Simpson found internal correlations between his various lists amounting to from 0.53 to 0.93 for his easy and to from 0.60 to 0.97 for his hard list. The test may, therefore, be regarded as possessing a good degree of reliability, particularly since Simpson's lists contained some words of unequal difficulty and are presumably less well

adapted for testing than the lists which have here been proposed for standard use.

(11) *Various correlations.* In the case of the women at Bedford Hills Reformatory, whose work was decidedly slow on the whole, the correlation between speed and accuracy was so high (0.83, P.E. .03, for those who needed no help in reading or writing) that the scoring was finally done in terms of accuracy alone.

Correlations determined by Simpson between the easy and the hard opposites and other tests were as follows: with the Ebbinghaus completion test 0.72 and 0.85, with memory for words 0.65 and 0.84, with the A-test 0.50 and 0.58, with memory for passages 0.50 and 0.70, with adding 0.56 and 0.70, respectively. These figures represent "estimated true correlations for people in general," as based upon raw correlations figured for his own adult *S*'s, corrected for attenuation and other probable sources of error.[5]

Thorndike found a very high correlation, 0.90, P.E. about .05, between the capacities of twins in this test.

Notes.—Special comparisons of different forms of controlled association have been made, among others, by Watt, Rusk and Winteler. The average association times reported by Watt are 1.364 sec. for part-whole, 1.454 for whole-part, 1.418 for co-ordinate, 1.548 for superordinate and 1.859 for subordinate relations. General agreement appears in the conclusions reached by Rusk, who lists the several varieties of association tested by him in the following order, passing from the easiest to the hardest: whole-part and part-whole, co-ordination, free concretes, superordination, subordination, free abstracts, causal. Winteler concludes that when superordinate, subordinate, co-ordinate, species-genus and opposite relations are tried with school children, the first takes the most and the last the least logical power.

The so-called 'B-method' of Ries, a test in which *S* is given a number of nouns representing causes and asked to name an-

[5]See the original article for the raw correlations and for correlations with other tests than those here cited. For Bonser's correlations, which refer to various special tests, see his monograph, p. 96.

other noun representing a related effect, yielded in his hands extraordinarily high correlations with estimated intelligence, 0.85, 0.86, 0.91 and 0.94 in different groups. The method is endorsed by Meumann (11, 432f.), who also reports excellent results achieved with it by Oksala, in Finland. Meumann declares that the capacity to seek out causes or effects affords a decisive index of degree of intelligence in children of from about 10 to 14 years, and perhaps older. We have made attempts to use this test in the educational laboratory, both at Cornell University and at the University of Illinois, and have encountered so many difficulties in the preparation of material, and especially in scoring, as to render the method unsatisfactory, even for mature college students.

Another relatively easy, though strictly controlled association test is that known as the backward-alphabet test. This has usually been conducted by asking *S* to name, or to write, as rapidly as possible, the letters that precede *f, k, s, p, w, l, e, r, d, o, v, j, n, t,* and *h.* For comparison, and to obtain a rough notion of *S*'s familiarity with the sequence of the alphabet in general, this test might be supplemented by another in which *S* was required to state the letters that follow another series of 15 letters.[6]

REFERENCES

(1) A. R. Abelson, Tests for mental deficiency in childhood. *The Child,* No. 3: 1912, 1-17. See also: The measurement of mental ability of 'backward' children. *BrJPs,* 4: 1911, 268-314.

(2) H. A. Aikins, E. L. Thorndike and Elizabeth Hubbell. Correlation among perceptive and associative processes. *PsR,* 9: 1902, 374-382.

(3) F. G. Bonser, The reasoning ability of children of the fourth, fifth and sixth school grades. *ColumbiaConEd,* No. 37. New York, 1910. Pp. 113.

(4) T. H. Briggs, Formal English grammar as a discipline. *Teachers College Record,* 14: 1913, 251-343.

(5) C. Burt and R. C. Moore, The mental differences between the sexes. *JEPd,* 1: 1912, 273-284, 355-388.

(6) D. F. Carpenter, Mental age tests. *JEdPs,* 4: 1913, 538-544.

(7) W. G. Chambers, Individual differences in grammar grade children. *JEdPs,* 1: 1910, 61-75.

(8) H. L. Hollingworth, The influence of caffein on mental and motor efficiency. *ArPs(c),* No. 22 (*ColumbiaConPhPs,* 20: No. 4). New York, 1912. Pp. 166.

(8a) H. L. Hollingworth, Variations in efficiency during the working day. *PsR,* 21: 1914, 473-491.

[6]On the use of this test, see Aikins, Thorndike and Hubbell.

(9) H. L. Hollingworth, Articulation and association. *JEdPs*, 6: 1915, 99-105.
(10) E. Jones, Some results of association tests among delinquent girls. *PsB*, 10: 1913, 78-79.
(11) E. Meumann, Vorlesungen zur Einführung in die exp. Pädagogik, 2d ed., Vol. II. Leipzig, 1913, especially 418-433.
(12) Naomi Norsworthy, The psychology of mentally deficient children. New York, 1906. Pp. 111. (Much of this material is also given in *JPsAsth*, 12: 1907-08, 3-17.)
(13) W. H. Pyle, The examination of school children. New York, 1913. Pp. 70.
(13a) W. H. Pyle, The mind of the negro child. *School and Society*, 1: 1915, 357-360.
(14) G. Ries, Beiträge zur Methodik der Intelligenzprüfung. *ZPs*, 56: 1910, 321-343.
(15) R. R. Rusk, Experiments on mental association in children. *BrJPs*, 3: 1910, 349-385.
(16) B. R. Simpson, Correlations of mental abilities. *ColumbiaCon Ed*, No. 53. New York, 1912. Pp. 122.
(17) Carrie R. Squire, Graded mental tests. *JEdPs*, 3: 1912, 363-380, 430-443, 493-506, especially 430-443 and 500-506.
(18) E. L. Thorndike, Measurements of twins. *ColumbiaConPhPs*, 1905. Pp. 64.
(19) E. L. Thorndike, Educational psychology, 2d ed., New York, 1910. Pp. 248. See also his Principles of teaching. New York, 1906, and his Introduction to the theory of social and mental measurements, 2d ed. New York, 1913. Pp. 277.
(20) H. J. Watt, Exp. Beiträge zu einer Theorie des Denkens. *ArGsPs*, 4: 1905, 289-436.
(20a) Jean Weidensall, The mentality of the criminal woman. To appear in *EdPsMon*.
(21) J. Winteler, Exp. Beiträge zu einer Begabungslehre. *EPd*, 2: 1-48, 147-247, especially 207-239.
(22) R. S. Woodworth and F. L. Wells, Association tests. *PsMon*, 13: 1911, Whole No. 57. Pp. 85.
(22a) Helen T. Woolley and Charlotte R. Fischer, Mental and physical measurements of working children. *PsMon*, No. 77, 18: 1914. Pp. 247. Especially 213-227.
(23) S. Wyatt, The quantitative investigation of higher mental processes. *BrJPs*, 6: 1913, 109-133.

TEST 34A

Analogies.[1]—In Test 33 the associations to be formed are left entirely to *S*'s choice, are unrestricted; in Test 34 the association is restricted to a single form of relationship throughout any one series. In the analogies test there exists restriction, but the kind of restriction varies from one stimulus to another

[1]The author is indebted to Professor D. Kennedy Fraser, of Cornell University, for the arrangement of this test.

within the series of terms. The kind of restriction, moreover, is not indicated to *S* in the instructions, but is supplied to him by the test material itself, and must be apprehended by him from that material. Each stimulus in the series consists of three terms; the first and second terms illustrate the relation in question; the third term is the first of a pair which are to stand in the same relation one to another as the first and second terms. *S*'s problem, then, is to find the appropriate fourth term. Because the relation varies from stimulus to stimulus, the test is sometimes referred to as the 'mixed relations test,' as, for example, by Woodworth and Wells (6), who say that it tests 'flexibility of mental performance' and also 'skill in handling associations.' Burt (2), from whom the term 'analogies' is borrowed, holds that the test involves "perception, implicit or explicit, of the relation and reconstruction of the analogous one by so-called relative suggestion." The test is recommended by these authors, as well as by Wyatt (7), and it needs little trial to show that it has many possibilities, particularly in view of the chance that it affords of constructing series of stimuli of varying difficulty.

MATERIALS.—(1) For individual tests: Split-second stop-watch. Three sets of 20 cards each, affording tests of three grades of difficulty. Sample set of 7 cards for preliminary trials. Prepared blank for registering times and incorrect answers. (2) For group tests or variant form of individual tests: stop-watch or special seconds-clock. Printed forms for each of the three sets of stimuli, provided with spaces for recording the responses.

The stimuli chosen for these series are taken from a large number of stimuli originally employed by Burt in work with the analogies test in England. They have been selected on the basis of fairly extensive trial with children and high-school and college students. If further materials are desired, *E* will find it more profitable to turn to the two lists of 20 stimuli each that are published by Woodworth and Wells (also reproduced by Briggs (1)) than to construct lists of his own, as only by actual trial can the feasibility of a given set of terms for use in this test be demonstrated.

METHOD.—Show *S* one of the sample cards and instruct him as follows: "On each of these cards there are three words, as on this one. As you see, there is relation between the first and

the second word. You are also given a third word, and I want you to find a fourth word which shall have the same relation to the third as the second has to the first. Work as rapidly as you can, and say the fourth word aloud as soon as you know what it should be. Thus, in this first card the fourth word is what? In the cards that follow the relation does not remain the same as this one, but varies from one card to another." If this explanation seems sufficient, proceed with the other sample cards, saying: "I will try these sample cards now to make sure that you understand."

Follow the instructions given in Test 34, part-whole test, including the use of the warning 'now,' the taking of times during the sample set, the correction of wrong responses, etc., save that it is unnecessary to record *S*'s response unless it be a word which is not provided for upon the prepared form. In ordinary testing *E* should pass to the next card whenever *S* is unable to give a response within 30 sec. Whether List A, List B, or List C, or some combination of them shall be used will depend upon *S*'s age and ability and upon the time at *E*'s disposal.

VARIATIONS OF METHOD.—Follow the suggestions for variant methods given for the part-whole test.

TREATMENT OF DATA.—This may be based, in general, upon the instructions already given for the part-whole test and for the opposites test.

The English investigators have attempted a somewhat finer scoring of quality of response than we have recommended. Thus, Burt scored 1 for each correct response, ⅔ for fair and ⅓ for poor responses, and 0 for omissions. Wyatt gave 4, 3, 2, 1 or 0 for responses grading from fully correct to omissions. This elaborate scoring is not needed for most of the terms in our lists, for in quite the majority of cases there is but one single correct response. In the other cases the use of the split-second watch in individual testing permits *E* to wait until the correct response is given and secure a direct measure of the time needed for this response, while the time at which the first (erroneous) response is made can also be put on record as an indirect measure of *S*'s general accuracy.

RESULTS.—(1) A general idea of the times that may be expected by the use of these three lists may be gained from the

results that have been obtained by their use in a limited number of cases (Table 71). The averages exceed the medians on account of the occasional very long association times which appear with nearly all *S*'s. The average time reported for their lists by Woodworth and Wells (about a dozen college and graduate students, using the method of exposing the entire list) is 3.14 sec., P.E. .13, with a range for individual averages of from 2.33 to 4.40 sec.

TABLE 71

Speed in Sec. for Correct Responses in the Analogies Test (Fraser)

		LIST A*		LIST B*		LIST C	
Group	Cases	Median	Aver.	Median	Aver.	Med an	Aver.
College	8	1.8	2.36	3.0	4.38	3.4	6.51
Adults	19	2.0	2.64	—	—	—	—
High-school Girls	30	2.4	3.16	—	—	—	—

*Lists A and B, at the time these figures were secured, contained 25 stimuli each. The omitted ones do not alter the conditions enough, however, to invalidate these records as norms.

More extended use of the Woodworth and Wells tests by W. V. Bingham, to whom I am indebted for advance figures from the results secured with 200 freshmen at Dartmouth College, has yielded the following percentile distribution for the analogies test (average time in sec. per response, based on two trials of 10 responses each) :

Poorest	10	20	30	40	50	60	70	80	90	Best
5.49	3.98	3.69	3.31	3.03	2.85	2.67	2.52	2.36	2.06	1.35

(2) *Dependence on age.* While sufficient data are lacking to present figures for various ages, there is a clear difference in the speed of *S*'s of grammar-school, high-school and college

standing when List A is used. Similarly, List B, and more especially List C, proves too difficult for younger *S*'s.

(3) *Dependence on sex.* Burt found an advantage of 15 per cent. in favor of the girls at the Wallasley School, Liverpool, *i. e.,* only 35 per cent. of the boys reached the median performance of girls. In other tests at the Holt Secondary School, however, the average performances of the two sexes were virtually identical. No sex difference appeared in our tests of Cornell University students.

(4) *Dependence on intelligence.* Wyatt, working with the time-limit method on groups, found that the analogies test afforded the highest correlations with intelligence of any of the tests he tried, save the completion test. His correlations amount to 0.62 in one and 0.80 in another group. Burt's tests at the Holt School gave a correlation between the results of analogies and intelligence of 0.50 in the individual test and 0.52 in the group test; his tests at the Wallasey School gave again a correlation of 0.50 (see Burt and Moore).

(5) *Reliability.* Burt's figures show that the analogies test possesses a good degree of reliability, as its coefficient of internal correlation figured in different trials 0.58, 0.71 and 0.92.

NOTES.—The analogies test appears to be better suited than other tests of association to bring out individual differences in quickness of adaptation to the task demanded. Thus, in the case of one high-school girl, the average association time for the first half of the list was 4.83 sec., for the second half only 2.19 sec. The inference that this *S* was naturally slow in adapting herself to new situations, but was able to work efficiently when once adapted, was afterward confirmed by the reports secured from her teachers of her performance in her school tasks, especially in geometry.

When *S*'s are tested by the standard method of securing the time for each response, it is often instructive to plot a rough frequency curve, with the second as a unit. A comparison of the distribution of the times for different *S*'s, as in the following sample, shows clearly individual differences in steadiness and consistency of performance as well as differences in general tendency toward fast or slow rates of mental activity.

Sample Distribution for Two College Students, List C

SECONDS	1	2	3	4	5	6	7	8	9	10+	MEDIAN	AVER.
Subject D	1	9	3	4	0	1	1	0	0	1	2.9	3.91
Subject G	0	4	3	1	3	2	2	1	1	3	5.6	5.67

It is also instructive to make notes of *S*'s general attitude toward the test, whether confident or hesitating, hurried or tranquil, etc., and to compare these attitudes with the quantitative results.

REFERENCES

(1) T. H. Briggs, Formal English grammar as a discipline. *Teachers College Record*, 14: 1913, 251-343.
(2) C. Burt, Experimental tests of higher mental processes and their relation to general intelligence. *JEPd*, 1: 1911, 93 112.
(3) C. Burt, The experimental study of general intelligence. *Child Study*, 4: 1911, 14-15.
(4) C. Burt and R. C. Moore, The mental differences between the sexes. *JEPd*, 1: 1912, 273-284, 355-388.
(5) R. S. Woodworth, The consciousness of relation. *Essays, philosophical and psychological, in honor of William James*, 1908, 485-507.
(6) R. S. Woodworth and F. L. Wells, Association tests. *PsMon*, 13: 1911, Whole Number 57. Pp. 85.
(7) S. Wyatt, The quantitative investigation of higher mental processes. *BrJPs*, 6: 1913, 109-133.

TEST 35

Controlled association: Computation.—The solution of simple arithmetical problems in addition, subtraction, multiplication, and division may be considered as essentially dependent upon the accuracy and rapidity with which the appropriate associative processes are executed. Computation is, therefore, a test of controlled association in which the restriction of the associative sequence is complete, in which only a single outcome is correct. But numerous subsidiary activities are, of course, involved. Thus, the solution of arithmetical problems with the aid of paper and pencil demands, besides associative activity, both visual perception and motor activity, while mental computation imposes an additional tax by necessitating the holding in mind of the problem itself and of the various steps in its solution.

Because of this implication of perception, movement, attention, retention, and perhaps other forms of mental activity, as well as simple associative activity, the computation test has been employed not merely for the special purpose of studying the nature and course of associative processes, but also for the more general purpose of investigating mental efficiency at large (*geistige Leistungsfähigkeit*). Oehrn, for example, who was one of the first to use computation as a mental test, sought to study individual differences in the nature of associative processes; Aikins, Thorndike, and Hubbell, Brown, Burt, Simpson, Hollingworth, and Krueger and Spearman to study the correlation of specific mental functions; Thorndike to determine the relative influence of heredity and environment upon mental efficiency; Reis to compare the ability of normal, paralytic, and hebephrenic children; Jones to investigate the effect of bodily posture, Vogt the effect of distraction, and Hollingworth the effect of caffein upon mental efficiency; Winch and Starch to investigate the transfer of special drill. But the commonest application of the computation test has been made in the formulation of the curve of mental efficiency, or the work-curve (*Arbeitskurve*), with special reference to the influence of practise, rest-pauses, exercise, and similar factors upon the mental efficiency of adults, and especially of children, during a school day. This use of the test is illustrated in the work of Arai, Bellei, Bischoff, Bolton, Burgerstein, Ebbinghaus, Friedrich, Heck, Heüman, Holmes, Katzen-Ellenbogen, Kafemann, Keller, Kemsies, Laser, Lindley, Marsh, Martin, Ordahl, Robinson, Schultze, Specht, Teljatnik, Thorndike, Wells, Weygandt, Winch and others.[1]

Addition, multiplication, and both in alternation, have been more popular forms of computation than subtraction or division. With all four forms varied types of problems have been used. These variations in the arrangement of the test naturally affect its outcome. The most important types of test are illustrated herewith. Beside the types that are shown, Winch

[1]A major portion of these studies are the direct or indirect developments of the special technique of the adding experiment as formulated by Kraepelin and his followers (see the various volumes of the *Psychologische Arbeiten* and Kraepelin's summary (27).

and others have employed miscellaneous arithmetical problems,[2] Reis had his *S*'s add mentally for 1 min. by 7's or by 12's, while Hollingworth has used a form of test, also recommended by Woodworth and Wells, in which a constant number is added (or subtracted) from a given list of numbers.

Examples of Material Used in Computation Tests

A	B	C	D	G
2	4	4 2 8 3 2 9 9 5 4 6 5 4 3 1 7	95799	93
6	1	7 9 2 9 3 8 3 8 2 6 5 5 1 3 9	86967	68
9	3	1 1 0 2 etc.	32687	41
7	5		84799	25
4	2		95976	52
8	6		34797	
9	4		97864	
5	3		98945	
1	4		87824	
7	6		68792	
5	1		79867	
2	5		88896	
6	4		97745	
3	2		39799	
5	1		48970	
1	2		89043	
3	3		67354	
9	6		54628	
2	5		91176	
1	3		90253	
3	6			
6	5			
9	1			
8	2			
3				
5				
2				
7				
6				
2				

H 6429364319483145762 7 + 3868272542358579185 8

I 982 — 469

J 64 — 27

K 28 × 8

L 363 × 6

M 47 × 89

N 948 × 579

O 7986 × 4523

R 4)799

P 428423995479253314325 × 4

Q 254)4659234(

[2]It is hardly necessary to allude to the development and use, particularly by Courtis, of special sets of tests for measuring the abilities of school children in the fundamental operations of arithmetic, the solving of arithmetical problems, copying figures and the like. Similar tests for algebra and geometry have been announced recently by other investigators. Tests of this variety are, of course, aimed at the determination of specific pedagogical attainments and differ, therefore, in scope and method from those here under consideration. For an account of the Courtis tests, with results of their application to a large group of school children, see Courtis (12).

EXPLANATORY

Addition

A. Vertical series of 1-place numbers, arranged to avoid repetitions and pairs adding to 10. *S*'s add continuously and drop back to units when each hundred is reached, or add by pairs, either orally or writing down the unit figure of each sum. The pairs are sometimes taken so that each digit is used twice, thus *S* adds 1 and 2, 2 and 3, 3 and 4, etc., and sometimes so that each digit is used once, as 1 and 2, 3 and 4, etc. Oehrn, Vogt, and others working under Kraepelin used columns of as many as 7,000 of such digits. Krueger and Spearman used 70, grouped by 10's as illustrated. Marsh used similar columns of 15 or 25 digits.

B. Vertical column of 24 1-place numbers, using 1 to 6 only. Used by Jones, who had *S* add aloud while he himself followed with a check list.

C. Horizontal series of 1-place pairs of digits. A modification of the Kraepelin 1-place series in order to make possible the examination of the accuracy of each addition. The unit figure of the sum is the only one recorded, as illustrated in the first four problems. Used by Schulze, and apparently also by Ebbinghaus and by Vogt.

D. Twenty 5-place numbers. Used by Thorndike.

E. Two 2-place numbers. Used by Teljatnik.

F. Two 3-place numbers. Used by Kemsies for mental addition.

G. Five 2-place numbers. Twenty such problems were given and 2 min. allowed for computation. Used by Thorndike and by Aikins, Thorndike and Hubbell. Four longer columns (25 numbers in each) are used in the Woodworth and Wells constant-increment test.

H. Two 20-place numbers. Used by Burgerstein, Laser, Friedrich, and Holmes. The last-named investigator published elaborate rules for the construction of these problems in such a way as to avoid the extension of errors in 'carrying.' She used 4 blanks with 16 such problems on each blank.

Subtraction

I. Two 3-place numbers. Used by Kemsies for mental subtraction.

J. Two 2-place numbers, to be written on the blackboard (Teljatnik).

Multiplication

K. Two-place multiplicand, 1-place multiplier. Used by Kemsies for mental computation, and by Ebbinghaus for written group tests.

L. Three-place multiplicand, 1-place multiplier. Used by Kemsies.

M. Two-place multiplicand, 2-place multiplier. Used by Keller, and by Marsh with the digits 1, 2, 5, and 9 excluded.

N. Three-place multiplicand, 3-place multiplier. Used by Keller for written, and by Thorndike for mental computation.

O. Four-place multiplicand and multiplier. Used by Thorndike and others both for written and for mental computation. The multiplicand is usually a combination of 6, 7, 8, and 9; the multiplier of 2, 3, 4, and 5.

P. Twenty-place multiplicand, 1-place multiplier. Used by Burgerstein, Laser, and Friedrich, with the restriction of the multiplier, in most tests, to 2, 3, 4, 5, or 6.

Division

Q. Three-place divisor, 7-place dividend. Four blanks of 10 problems each were used by Bellei for an hour's work.

R. One-place divisor, 3-place dividend. Used by Kemsies for mental computation.

There are certain advantages and certain disadvantages in each of these forms of material. In general, *E* must select that form of test that best suits the conditions under which he works.

MATERIALS.—Stop-watch, preferably split-second. Printed forms, containing problems in addition and multiplication. For group tests, the special seconds clock.

Five forms have been prepared for this test: others may be prepared by *E* as desired.

A. Addition test: several thousand digits in vertical columns with a line separating each 10 digits, after Model A. This form may be used with children or with adults, and either for short series or for continuous adding, after the Kraepelin method, after the plan of Krueger and Spearman, or after the method of adding pairs.

B. Addition test with 36 problems, patterned after Model G, but containing 10, in place of 5 numbers each. This can be used also for tests in which a constant number is added or subtracted.

C. Addition test, patterned after Model C (Schulze's method), and specially recommended for younger *S*'s.

D. Addition test, patterned after Holmes, Model H, and virtually identical with the material used by Burgerstein, Laser, and Friedrich.

E. Multiplication test, after Model P, as used by Burgerstein, Laser, and Friedrich.

METHOD.—(1) *General determination of S's ability* may be carried on with any one of the forms. The following general principles should be kept in mind: (*a*) Individual tests are usually more satisfactory than group tests. (*b*) Any computation work that is so easy that the mental operations can proceed as fast as the results can be written (as Form C for adults) would better be given individually and arranged so that *S* may announce the results orally and *E* check them off upon a prepared key; and in general, care must be taken that the recording of results shall not fall to *S* unless it is certain that his associations will neither be delayed nor disturbed by the process of recording. (*c*) Group tests with competent *S*'s may be most satisfactorily carried on by the work-limit method with the aid of the seconds clock. (*d*) Group tests by the time-limit method should, as a rule, be terminated at such a time that the fastest *S* in the group can no more than complete the task.

In accordance with these general principles, Forms B, D and E will be found adapted for group tests or for individual tests

with the recording of the figuring done by *S* himself. *E* can take the time for performing any specified number of the problems, or he may also, especially by using a split-second stopwatch, secure the exact time for solving each problem without interrupting *S*'s work until the entire test form is finished.

For the *constant-increment test* *S* is given Form B, printed side down. He is instructed at the signal 'now' to add a specified number to each number found in the columns when he turns over the form. The numbers commonly used have been either 4 or 17. The test may be repeated with other increments, and these may be so chosen as to secure wide variations in difficulty, as by assigning easy constants, like 1 or 2, or more difficult ones than 17. Again, the assignment may be to subtract a given number.[2]

Form C affords a particularly good test of skill and accuracy in the addition of units, especially when conducted orally. Record the time for each row horizontally. Adults will make but few errors, and these they may be allowed to pass over or to correct, whichever way they may prefer.

Form A may be given by a variety of methods. In particular, *S* may write down the sums for each section of 10 digits and *E* record the time for each section until one page of the material has been covered: or *S* may add orally by pairs while *E* watches for errors upon a prepared check sheet, and also notes upon it the place reached by *S* at given time-intervals, as at each minute or each half-minute; the adding in this case may be done by either of the methods of grouping the pairs mentioned above (Explanatory, Addition, A).

(2) For those who wish to arrange an experiment for the *special determination of S's susceptibility to practise, fatigue, etc.*, some suggestions may be found in the following development of the method of Kraepelin illustrated in the work of Specht and of Bischoff. To carry out this experiment fully, *S* adds by pairs, 10 min. per day, on each of 12 successive days.[3]

[2]For timing work by columns a convenient arrangement is to cut them out of the form and paste them singly upon small stiff cards.

[3]It would seem possible to condense this time, either by taking fewer days or by adding during several sittings on a given day, though it is impossible to predict whether the results would then be comparable to those reported below from Specht and Bischoff.

The pairs are added by the 1 and 2, 2 and 3 method and the unit figure only of each sum is written down by *S*.[4] A bell-stroke or other signal is given at the end of each minute, and *S* marks by a horizontal stroke the point he has reached at the signal. On the 1st, 3d, 5th, 7th, 9th and 11th days there is introduced between the 5th and the 6th minute of the adding a rest-pause of 5 min.: on even-numbered days *S* adds directly through the 10 min. without pause.

In carrying out this special form of addition test, or in fact, in carrying out any test which is designed to measure efficiency under various conditions—different periods of the day, after recesses, after gymnastics, after eating, etc.—it is evident that *E* must bear in mind the possibility that a number of different factors may enter to affect the performance, and that to measure any single factor, like fatigue, the influence of these other factors must be excluded or allowed for. The most serious of these disturbing factors are practise, excitement, ennui and carelessness.

A common method for cancelling out practise is to divide *S*'s into two equivalent groups on the basis of a preliminary test, and to administer one set of problems early to the first, and late to the second group (if, for instance, fatigue is to be investigated), the other set late to the first, and early to the second group.

In studying the work curve, some *E*'s have used computation both as the test and as the work to induce fatigue, practise, etc.; others have used computation as a test of efficiency, but have allowed *S* to follow in the main the regular work of the school session. In the first procedure, computation (usually addition) is pursued more or less continuously for an hour, or even for several hours; in the second procedure, the computation itself occupies but a short time, relatively, say from 1 to 10 min., and is repeated at intervals of an hour or more, while *S* meantime takes up his regular tasks, indulges in physical activity, or rests, as *E* may direct.

In illustration, Vogt, Oehrn, and other disciples of Kraepelin, have kept their *S*'s adding continuously for several hours; Holmes used 4 periods of adding of 9 min. each, with 4-min. rest-pauses, Burgerstein 4 periods of 10 min. each, with 5-min. pauses. Typical illustrations of the second procedure are supplied by the investigations of Laser and of Ebbinghaus, who introduced 10-min. computation tests at the beginning of the school day and once an hour thereafter, and also by the studies of Heck and of Robinson. Ebbinghaus is inclined, however, to recommend 5-min. tests as being equally serviceable for the determination of efficiency and less likely to develop ennui and carelessness. Offner (35, p. 48) favors short tests for similar reasons and also for the partial avoidance of the practise-error.

Treatment of Data.—Computation tests yield two measures of efficiency—speed (or quantity of work) and accuracy (or

[4] It would seem to the author much preferable to use oral adding, but here, again, it is not possible to predict what effect such an alteration of method might have upon the results.

quality of work). Many investigators, particularly when examining the effect of practise, fatigue and similar factors upon performance, have found it best to keep the two measures separate. Some investigators, like Teljatnik, have considered quality of work only; more often, qualitative differences, being relatively small, have been disregarded and performance has been ranked by speed of work only. The combining of speed and accuracy into a single score representing net efficiency may be attempted by some of the methods proposed in Test 26. Or, an arbitrary penalty may be contrived for each error and the time consumed may be increased by these penalties. Thus, Simpson, who used material like our Example G, computed the final score of his *S*'s by adding to their actual time 10 sec. for each error. An *S* who added 10 examples in 55 sec. and got seven answers right and three wrong would then be given a final score of 85 sec.

Quantity of work is indicated by elapsed time when using the individual method, and by the number of problems solved (sometimes by the number of figures written in the results) in the time-limit method.

Quality of work is generally regarded as directly proportional to the percentage of correct solutions. Inaccuracy is most often taken in terms of the number of errors committed, less often in terms of the number of errors plus the number of corrections made by *S*. The simplest, but the least desirable way to compute errors is to score one error for every wrong figure in the result. In the case of certain problems, however, a single error in computation may affect more than one figure in the result.[5] For reliable results, these complex errors must be examined and the score adjusted to indicate exactly the number of real errors of computation.

For the special experiment patterned after Specht and Bischoff more elaborate treatment of data is called for. (1) The gain in sums added the 6th min. as compared with the 5th min., in its relation to the sums added the 5th min. (*i. e.*, the per cent. of gain) is computed both for all the days with pause

[5]The problems in Form D (Example II) are intentionally arranged to reduce this error.

and for all the days without pause, and the difference between these two relative gains is found. (2) The number of sums added in the first five and in the second five minutes, both on days with and on days without pause, is treated in the same manner. (3) The difference between the sums added on the 2d and the 6th minute on days with pause, taken as a per cent. of gain over the 2d minute, forms the coefficient of practise. (4) To find the coefficient of fatigue

let F = the required coefficient of fatigue,
P = the coefficient of practise,
A = the sums added the first 5 min. without pause,
B = the sums added the second 5 min. without pause, and
b = the sums theoretically added the second 5 min. under practise, but not under fatigue.

Then $b = \frac{A \times (100 + P)}{100}$ and $F = \frac{100 \times (b - B)}{b}$.

Thus, if P = 8.8, A = 1226, B = 1141, then b = 1333.9 and F = 14.46. (5) The difference between the sums added in the 2d min. and in the 10th min. of days without pause, taken as a ratio to the 2d min., affords another, and in some respects, a better index of fatigue. (6) The total number of sums added the 1st 5 min. of all days gives T (total performance), which affords an approximate notion of *S*'s ability to add. *T* is also made the denominator of a fraction, the numerator of which is the total number of additions made the very first 5 min. The fraction gives some indication of *S*'s susceptibility to practise. (7) Another index of susceptibility to practise is secured by taking the average of the gains in the first 5 min. from day to day as against the first day and figuring the difference as a ratio to the first 5 min. (1st day). Characteristic results for all these values are given below.

RESULTS.—(1) Woodworth and Wells report the average time of college students in the Kraepelin form of adding as 107.2 sec., range 65 to 164 sec. The same authors report for the constant increment test (one column of 25 numbers) adding 4: average 33.9, range 24 to 49 sec.; subtracting 4, average 41.1, range 25 to 67 sec.; adding 17, average 97.4, range 62 to 158

sec., with an average of 2.4 errors in the last form of test. These figures are based upon a very limited number of *S*'s (7 to 10). In the author's laboratory, tests with college students in adding 50 sections of 10 digits on the Kraepelin form have yielded individual averages per section of from 5.8 to 13.7 sec., while the number of correct sections has ranged from 34 to 46.

(2) In all computation tests, and particularly in those embodying mental multiplication, there are marked *individual differences* in speed and accuracy, even among *S*'s of the same age and same school grade. Thus, Schulze's best pupil added more than 5 times as fast as the slowest pupil in the same class.

(3) *Dependence on sex.* There is evidently no decided sex difference in computation, since the results of various investigators are conflicting. Burt found girls slightly slower in multiplication; Burt and Moore reckon that 65 per cent. of boys exceed the 50 per cent. record of girls in adding and 63 per cent. exceed the 50 per cent. record of girls in multiplication. On the other hand, Courtis' New York results show that girls are slightly better than boys in the fundamental operations of arithmetic. Again, Miss Holmes found girls slightly better than boys, and the conclusion of Fox and Thorndike is that the girls in the high school they studied were about 5 per cent. better than boys, though here there may have been a better grade of girls selected by the school. In the solving of arithmetical problems, however, where something more than knowledge of the fundamental operations is involved, the work of Courtis and of Thorndike (52) shows a superiority of boys amounting to an excess of some 10 per cent. in the distribution above the median of the girls. "Roughly, boys are about half as far ahead of the girls in the same grade as they are of the boys in the preceding grade." Heck found that boys fell off more in quality of work in the afternoon session than did girls (4.25 vs. 1.96 per cent.)—a result possibly due to a greater carelessness on the part of the boys.

(4) *Dependence on school grade.* When sufficiently large groups are compared, there is, of course, a perceptible differ-

ence between the performance of one grade and that of the grades above or below it, but this difference is small in comparison with the range of variation within any grade, and may, on that account, disappear when small groups are compared. Thus, the curves of distribution in adding reported by Chambers for 22 seventh and 22 eighth grade pupils cannot be distinguished, while Courtis generalizes results for his multiplication test by saying that "35 per cent. of any grade membership will exceed the average score of the next higher grade: also, that 35 per cent. of the grade membership will fall below the average of the next lower grade" (12, p. 450).

(5) *Dependence on practise.* All investigators agree that practise produces a considerable improvement in all forms of computation, despite the fact that the associative connections concerned have been long established and often used. Hollingworth, who used the constant-increment test (adding 17 to 50 2-place numbers) found that, even after 35 preliminary trials, one of his groups reduced their average time from 102.7 to 61.2 sec. during 17 further trials, a reduction of some 40 per cent. Similarly, the 19 university students reported by Thorndike, who added daily for a week 48 columns of ten numbers, effected a median reduction in time of about 31 per cent., and in accuracy of about 29 per cent., although the total amount of time spent in the work was only about one hour for each *S*. Not all these *S*'s showed such practise effects; for one or two there was no improvement, while one improved as much as 50 per cent. It is worth noting that practise-improvement is shown by those who stand high at the beginning of the work as well as by those who stand low then. The same result has been found also in tests of 29 boys in a New York City 4th grade school, where, according to Donovan and Thorndike, those most efficient at the beginning gained on the average as much or more (in gross gains) as did those least efficient at the beginning. Wells' tests of adults (56) lead to a similar conclusion. On the whole, however, practise in adding tends to reduce somewhat the initial differences between the *S*'s, whereas practise in mental multiplication seems not to affect much the relative differences between *S*'s, from which Thorndike con-

cludes that the abilities demanded in mental multiplication are more dependent upon original capacity than are those demanded in adding. Mrs. Ordahl found that practise in mental multiplication produced a decided gain in speed without much improvement in accuracy, and she believes that the improvement in this operation resides more in the methods of handling the task than in the facilitation of the numerical associations themselves.[6]

The question of the transfer of practise-effects in computation to other forms of mental activity has been studied by Winch and by Starch. Winch was unable to decide whether special drill in computation produced an increase of skill in solving arithmetical problems; there appeared to be a transfer in some of the classes, but not in others. Starch found that a 14-day drill in mental multiplication developed an improvement of from 20 to 40 per cent. in other arithmetical operations, but had little effect upon auditory memory span.

(6) *Dependence on intelligence.* Burt tested English school children, aged 12, to determine the number of additions or multiplications correctly made in 10 min., and found a correlation with intelligence of 0.25 in addition and of 0.41 in multiplication. Brown's results for a group of 39 girls, aged 11 to 12, show no correlation between school grades and speed or accuracy of adding and a correlation of only 0.10 between speed of adding and estimated general intelligence: his results for another group of 40 boys of the same age show correlations of 0.28 between speed of adding and school marks, of 0.24 between speed of adding and estimated intelligence, and of 0.11 between accuracy of adding and marks, with absence of correlation between accuracy and estimated intelligence. Simpson used adding in his study of two sharply-contrasted groups of adults; the test separated the groups fairly clearly—only 10 per cent. of the poor group reached the median performance of the good group. Within the good group the results of the adding test correlated by 0.72 with estimated intelligence.

(7) *Reliability.* Save for Burt's figures (0.50 for adding and 0.55 for multiplication), the internal correlations for com-

[6] Consult her article for detailed tables and introspective reports.

putation tests show a good degree of reliability. Krueger and Spearman, for instance, obtained a reliability of 0.76, Simpson of 0.76 for his good group, 0.90 for his poor group, 0.91 for both together. Brown, who worked with several different groups, reckons the reliability for speed of adding at 0.68 to 0.98 and for accuracy of adding at 0.30 when one application of the test is made, and reliability for speed at 0.81 to 0.99 and for accuracy at 0.36 to 0.74 when the scores represent amalgamated results of two applications.

(8) *Correlations.* Aikins, Thorndike, and Hubbell compared efficiency in adding with efficiency in the other 'association' tests (misspelled words, cancellation of two letters, and opposites), and (by a special method of estimating the index) found the quality of work in adding and quantity of work in associating correlated to a degree of 50 per cent. in 8th-grade, and 20 per cent. in 5th-grade pupils, and net efficiency in adding and net efficiency in associating correlated to a degree of 48 per cent. On the other hand, the percentage of error in adding and in the other association tests exhibited no correlation or one of but slight degree.

Thorndike's study of mental resemblances in twins (47) showed a much higher correlation of ability in computation between twins than between siblings; thus, twins aged 9-11 years revealed a correlation of 0.90 in adding, and 0.91 in multiplication, and twins aged 12-14 years a correlation of 0.54 in adding and 0.69 in multiplication: taken collectively, the index of correlation amounted to 0.75 for the adding, and 0.84 for the multiplication test.

Fox and Thorndike found that ability to add correlated to a fairly high degree, 0.75, with ability to multiply, but only to a small degree, 0.20 to 0.44, with ability to solve fractions or to perform other arithmetical problems. They conclude that "ability in arithmetic is thus but an abstract name for a number of partially independent abilities."

These results do not agree well with those reached by Winch in his two studies of the transfer of drill in numerical accuracy, since he found high correlations (0.68, 0.69 and 0.74) between accuracy in computation and in arithmetical reasoning.

Burris found that school grades in algebra and in geometry, as recorded in 19 representative high schools, showed, for nearly 1000 pupils, a correlation of 0.45.

Simpson publishes the following "estimated true correlations for people in general" with efficiency in adding: Ebbinghaus completion test 0.65, hard opposites 0.70, easy opposites 0.56, a-test 0.58, memory for passages 0.42, estimation of lengths 0.00. The extended series of intercorrelations found by Brown, which are in general much lower than those of Simpson, will be found reproduced in Simpson (41, 107f.) as well as in the original text (5, 309-313, 316).

Krueger and Spearman found a good degree of correlation between adding and pitch discrimination (raw correlation 0.67, 'corrected' correlation 0.68, 'completed' correlation 0.80) and between adding and the Ebbinghaus test (raw correlation 0.79, 'corrected' correlation 0.68, 'completed' correlation 0.93).

Hollingworth (20) has studied the effect of a long series of trials (over 200) upon the correlations between adding and various tests; the following are typical results:

	TAPPING	CO-ORDINATION	DISCRIM. REACTION	COLOR NAMING	OPPOSITES
1st trial	.45	.21	.23	.26	.23
205th trial	.57	.16	.15	.76	.76

(9) *Relation of speed and accuracy.* While it is doubtless true that, for a given individual working under constant conditions, an increase of speed tends to produce an increase of errors, it is equally true that under actual working conditions a given individual may show an increase of speed coupled with a decrease in number of errors. When individuals are compared, it is found that the faster *S*'s are, on the whole, also the more accurate *S*'s. In six of his groups Brown found correlations between speed and accuracy of adding ranging from 0.13 to 0.43, P.E.'s from .07 to .12. With small groups of college students I have obtained similar positive correlations of 0.19 in the case of adding and as high as 0.86 for mental multiplication.

(10) *Mental defectives.* Reis found that with paralytics and hebephrenics the average performance was less and the variability greater than with normals. Similarly, Specht says that, though there appear decided individual differences in fatiguableness as shown by the adding test applied to normal individuals, a still greater amount of fatiguableness appears when it is applied to patients in an insane hospital selected for their tendency toward easy fatiguableness in general. I have combined the data furnished by Bischoff for 12 normal *S*'s and by Specht for 17 normal and 6 insane *S*'s when tested by their special form of the Kraepelin addition test and scored according to the directions given above for that experiment. The results are given in Table 72. The differences between the two groups are readily obvious in Factors 8 and 9, which reveal tendency toward fatigue, and in Factors 10 and 11, which show the total amount of work done. Analogous results have been reported by Katzen-Ellenbogen, who concludes that "the average curve of epileptics is decidedly different from the normal [curve] and characteristic of epilepsy."

TABLE 72

Average Scores of 29 Normal and 6 Insane Subjects in the Addition Test (After Specht and Bischoff).

		Normals.	Patients.
(1)	Per cent. gain 6th over 5th min., with pause..	14.7	20.6
(2)	Per cent. gain 6th over 5th min., no pause....	—1.0	—3.9
(3)	Difference between (1) and (2).....	15.7	24.5
(4)	Per cent. gain 2d over 1st 5 min., with pause..	4.4	—0.3
(5)	Per cent. gain 2d over 1st 5 min., no pause....	—5.0	—11.4
(6)	Difference between (4) and (5).....	9.9	11.1
(7)	Coefficient of practise.....	9.9	13.3
(8)	Coefficient of fatigue.....	—12.6	—21.3
(9)	Per cent. gain 10th over 2d min., no pause.....	—5.6	—14.3
(10)	Additions 1st 5 min. in 1st trial.....	172.3	87.5
(11)	Total additions 1st 5 min. of all trials.....	3406.8	1157.6
(12)	Progress of practise.....	11.1	2.0

(11) *Miscellaneous influences.* Hollingworth (19) used adding among other tests in his study of the *effects of caffein* and found that this drug produces pronounced stimulation in the processes of adding.

Posture was found by Jones to affect the speed of adding; both children and college students could add somewhat faster

(approximately 3 to 8 per cent.) with the body in a horizontal, than with the body in a vertical position.

The *effect of distraction by concomitant* activities, *e. g.*, the reciting of a poem, was found by Vogt to reduce very materially (58.5 per cent.) the number of additions made by the continuous (Kraepelin) method, but to have relatively little effect upon the simpler process of adding pairs of digits.

(12) *Fatigue and other factors of the work curve.* The use of computation tests to determine general mental efficiency at different hours of the day, with special reference to the performance of school children under classroom conditions, represents a special form of experiment that oversteps the boundaries of mental tests in their diagnostic use. In what follows, therefore, I have limited the treatment to presenting typical results and to pointing out certain important general principles that have been established in this field.[7]

(*a*) *General analysis of the work curve.* In other tests (especially Nos. 10 and 26) we have had occasion to refer to the fact that attempts to isolate fatigue from other influences affecting a curve of work are rendered difficult because of the presence of other complicating factors. Of these, practise is perhaps the most obvious and influential. Investigators have added, however, numerous other factors, such as recuperation, adaptation, momentum, swing, or fitness for work, warming-up, and spurts of various kinds. Extended accounts of these factors will be found in Meumann (33, II, 8ff. and elsewhere), Schulze (40, 320ff.), and particularly in the writings of Kraepelin (26, 27) and his students. However patent and plausible these factors may appear from observation of our daily activities, it seems probable that they have sometimes been invoked in explanation of work curves when actual demonstration of their existence is difficult, if not impossible.[8] In work curves obtained from school children it is certain that loss of interest,

[7] For a comprehensive critique of the experimental literature upon work and fatigue the reader is referred to Thorndike (53). In my translation of Offner (35) there will also be found a more general discussion of the whole topic of mental fatigue.

[8] Thorndike is especially severe in his criticism of the numerous lesser factors exploited by the Kraepelinian school.

or ennui, and resultant carelessness complicate the performance seriously, and are often mistaken for the effects of true fatigue, in the sense of actual inability to work at a sustained level of efficiency.

(*b*) *Individual differences in the work curve.* We have noted the presence of clear-cut individual differences in the speed and accuracy of computation; there are also individual differences in the course of the performance. Thus, both Kemsies and Keller conclude that mass results should be subjected to scrutiny to detect individual curves of performance if reliable information is to be secured concerning fatigue and overpressure in the schools. The recent work of Miss Martyn (32), similarly, has shown that the introduction of a rest pause may be favorable to some *S*'s and unfavorable to others, and also that the effect of fatigue may be met and masked in some *S*'s by the presence in them of a strong permanent 'set' for accurate work. "We may conclude," she says, "that fatigue cannot be invariably estimated by diminution either in speed or in accuracy of work, since habit and method of working bear an important relation to its manifestations" (32, p. 434). Again, the results obtained by Miss Arai and confirmed by numerous investigators make it fairly certain that the most competent workers are the ones least affected by fatigue.

These individual differences in susceptibility to fatigue have tempted some investigators to sort *S*'s into certain groups or 'types' of workers. If by 'types' is implied that individuals can be sorted into 'water-tight compartments,' the hypothesis must be regarded as of doubtful utility. Illustrations are seen in the work of Kemsies, who distinguishes between persistent workers who fatigue slowly and profit much by practise and feeble workers who fatigue quickly and do not profit much by practise. Meumann's own investigations lead him (33: vol. 2, pp. 10-11) to posit three types of workers (quantitatively regarded): the first type attains maximal efficiency at the start and thence decreases with many fluctuations: the second attains maximal efficiency only after an interval (of a length depending upon the kind of work); the third attains maximal efficiency only after a long period, perhaps several hours, of

work. The first type, then, is characterized by rapid adaptation and rapid fatigue, the second by slower adaptation and slower fatigue, the third by very slow adaptation and very great resistance to fatigue. The third type, he thinks, is probably more common in adult males, the first in women and children.

(*c*) *The work curve for continued computation.* 1. Work without interruption. Oehrn found that when adults added continuously for 2 hours or more, maximal speed was attained on the average at about 28 min. from the start. Schulze finds, however, that with school girls aged 12.5 years, signs of fatigue appear even in the first 5 min. The total number of additions made per minute by 37 girls was 1850, 1871, 1863, 1785, and 1772 for the 1st to the 5th minute, respectively.

Schulze's results with the same pupils for longer periods (50 min. without pause) show a progressive decrease both of quantity and of quality of work (Table 73). These figures, which are selected from the 6th of a series of experiments, are based upon the very easy process of adding two 1-place digits, so that practise has relatively little effect, but fatigue diminishes efficiency.

TABLE 73

Efficiency in Addition: Five 10-Minute Periods (Schulze)

PERIOD OF TEN MINUTES	TOTAL NUMBER OF ADDITIONS	PERCENTAGE OF DECREASE OF QUANTITY OVER THE PREVIOUS PERIOD	PERCENTAGE OF DECREASE OF QUALITY OVER THE PREVIOUS PERIOD
I.	17,740	—	—
II.	16,726	5.7	.09
III.	15,855	5.2	.03
IV.	15,485	2.3	.17
V.	15,134	2.3	.01

The effect of continuous work upon a very difficult task depends upon the degree of practise previously attained, the actual length of the work and the general condition of *S* when it is begun. Thus, Thorndike (51) induced 72 college students to multiply 3-place numbers mentally for about two hours,

with the net result that the work improved somewhat both in speed and accuracy; nevertheless, a rest of 30 min. effected an increase of about 5 per cent. in speed and a rest over night a still further increase in speed of about 7 per cent. But when the same investigator had 16 *S*'s mentally multiply a 3-place by a 2-place number continuously for from 3 to 8 hours, or (with pauses for meals) from 4 to 12 hours, only 3 *S*'s did as well at the end of their work period as when they had rested; the results showed, as might be expected, a compound of gradually lessening practise and gradually increasing fatigue. Miss Arai, who mentally multiplied 4-place numbers for 11 or 12 hours at a stretch after practise-effects had been largely eliminated, found that the time needed to work such examples was practically doubled at the end of eleven hours.[3]

2. Work with interruptions. When repeated computation tests are made within an hour, the usual result is a progressive increase in the quantity, but a progressive decrease in the quality of the work. Burgerstein's figures (Table 74) furnish a typical example of the results for four 10-min. periods with 5 min. rest-intervals between periods.

TABLE 74

Efficiency in Addition and Multiplication within an Hour (Burgerstein)

PERIOD	NUMBER OF FIGURES IN RESULTS	NUMBER OF ERRORS	PERCENTAGE OF ERROR
I.	28,267	851	3.01
II.	32,477	1292	3.98
III.	35,443	2011	5.67
IV.	39,450	2360	5.98

Miss Holmes' results are similar, though, on account of computing errors of a different plan ('serial' errors counting but

[3] In a test conducted under my direction and as yet unpublished, Mr. Painter, after preliminary trials to remove most of the effect of practise, worked at difficult mental multiplication, beginning late in the evening after a day's university work and continuing until the task became impossible. The cessation of work was not gradual (with ability, for example, to multiply 2-place numbers when 4-place were impossible), but appeared as a sudden collapse such that mental work of any sort was quite impossible.

as one error), her percentage of error averaged but 1.3, as against Burgerstein's 3.[10]

The common interpretation of results like Burgerstein's has been that practise increases the speed of the work, while fatigue increases its inaccuracy. But Ebbinghaus (14, pp. 406f.) denies that practise could produce such marked increase of speed, and ascribes both the increase of speed and the decrease of accuracy primarily to increased haste and carelessness.

(*d*) *Effect of rest-pauses.* When, either from ennui or fatigue, efficiency tends to decline, a period of rest generally exerts a favorable effect. With school children, as would be expected, such a pause is favorable even after relatively short work, as is illustrated by the data of Table 75, which are derived by Burgerstein from Schulze. The effect of rest upon efficiency in mental multiplication after some two hours' work has already been mentioned with reference to experiments with college students. The tests made by Friedrich upon 10-year-old pupils and by Kraepelin upon adults (26, pp. 16-17) furnish similar evidence of the effect of rest-pauses.

TABLE 75

Additions per Pupil, with and without a Rest-Pause (Burgerstein-Schulze)

	FIRST 25 MINUTES	REST-PAUSE	SECOND 25 MINUTES
First test	1067	5 min.	1088
Second test	1146	None	1042

(*e*) *Efficiency at different periods of the day.* Typical instances of the use of computation as a test for the fatigue-effects of the regular school program are afforded by the experiments of Friedrich, of Laser, and of Ebbinghaus. This method has been adopted in part to avoid the entrance of ennui and carelessness previously mentioned.

[10] Miss Holmes' analysis of the errors showed that their increase during the hour was due primarily to increased inaccuracy in associative processes, rather than to increased frequency of 'slips of the pen.' In general, errors of transcription were about one-third as numerous as errors of association.

Laser's tests, at hourly intervals, of 226 pupils (aged 9-13 years) in a Königsberg *Bürgerschule* are summarized in Table 76. Inspection shows that, save for the 5th period, the outcome is the same as that of the tests for an hour's time by Burgerstein, viz.: a progressive increase in speed and decrease in accuracy of computation.

TABLE 76

Efficiency in Computation within a School Session (Laser)

TEST AFTER SCHOOL PERIOD	TOTAL NUMBER OF FIGURES ADDED	TOTAL NUMBER OF ERRORS	PERCENTAGE OF ERROR
I	34,900	1147	3.28
II	40,661	1460	3.59
III	43,124	1713	3.79
IV	43,999	1796	4.08
V	45,890	1668	3.63

Ebbinghaus, who sought to determine the desirability or undesirability of a 5-hour continuous school session in a *Gymnasium* and higher girls' school at Breslau, obtained results identical with those of Laser so far as the qualitative aspects are concerned, but differing somewhat as regards the quantitative aspects, more particularly in that speed of computation reached a maximum at the close of the 2d school period, to remain thereafter almost constant or to fall off slightly toward the close of the session. Friedrich's results lead him to advise lighter work in the afternoon session. Bellei found that boys and girls aged 12 solved problems in division more slowly and less accurately in the afternoon than in the morning. Marsh tested but a few individuals, so that it is probably unsafe to make inductions from his data, which seemed to indicate a greater efficiency in adding at noon than later in the day, and in multiplication at between 1.30 and 3 p. m. than at 6 or at 10.30 p. m. Miss Martin had 6 *S*'s add for 15 min. at 10, 12 and 4 o'clock, with the result that slightly more sums were completed at 12 and somewhat fewer at 4 than at 10; the differences are, however, inside the probable error. The work of the first 5 min. was relatively poorer in the afternoon, due, she

thinks, to a later entrance of *Anregung* at that time. The most authoritative laboratory study of efficiency in calculation at different periods of the day, however, is that of Hollingworth (20a), who had opportunity during his experiments upon the effects of caffein to watch the daily curves of *S*'s whose work was done under exceptionally good conditions as regards elimination of practise error. In the use of the constant-increment test (adding 17 to 50 2-place numbers) at 8, 10, 12, 3 and 5.30 o'clock there appeared progressive fatigue amounting to about a 2 per cent. lengthening of the time at each trial, with a total lengthening of 7.50 per cent. in the case of 5 women and 10.5 per cent. in the case of 5 men. In further use of the same test in a more intensive experiment (15 trials between 10.30 A. M. and 10.30 P. M.) there appeared, again, a lengthening of about 10 per cent. toward the end of the day.

Heck tested 1153 New York school children (18) and 573 Lynchburg, Va., children (17) with a modification of the Courtis tests for the fundamental arithmetical operations. The New York tests lasted 10 min., those at Lynchburg 25 min., and they were distributed over various periods of the school session, particularly at 9, 11, 1 and 2.30 o'clock. The general result was an increase in quantity and a decrease in quality toward the close of the day; at New York, for instance, quantity increased by 1.57, 1.64 and 2.36 per cent. in the 2d, 3d and 4th periods, while quality decreased by 1.51, 1.41 and 2.28 per cent. in the corresponding periods. These differences are so slight as to be pedagogically negligible, in the opinion of Heck. The inferior quality of the later periods is, he thinks, more likely a sign of lessened interest than of consumption of energy or any sort of fatigue-poisoning. Rather elaborate tests with computation and other forms of school work by Robinson in South Carolina show in general little evidence of actual loss of ability toward the close of the school session.[11] The same conclusion has been reached by Thorndike (45) from schoolroom tests at Cleveland, Ohio, and Scranton, Pa. He emphasizes the statement that

[11]Consult the original for a discussion of the effects of recesses, lunches, gymnastics, singing, special incitement, etc., upon performance in such tests. The main conclusions are also summarized in *JEdPs*, 3: 1912, 593-595.

"incompetence, mental fatigue, does not come in regular proportion to the work done," that feelings of fatigue are not measures of mental inability, that disinclination to work does not signify inability to work. It may be questioned, however, whether this demonstration that pupils *can* work nearly as well at the end of school session as at its beginning is equivalent, as some writers have thought, to a demonstration that they should be expected to work as well at the later periods.

A special study of *fatigue in evening schools* by Winch leads him to the conclusion "that evening work is comparatively unprofitable, and that a short time in class in the evening is sufficient, *plus* the labors of the day, to induce a low condition of mental energy."

NOTES.—Those who have used computation tests have not sought, as a rule, to examine the mental processes involved in them. Oehrn, however, calls attention to the fact that practise in adding (by the Kraepelin method) tends to induce quasi-automatic addition. This circumstance, taken in conjunction with the relatively small correlations between different forms of computation themselves, and between them and other abilities, including general intelligence, lends countenance to Wells' objection (55) to accepting the computation test, without further qualification, as a measure of general mental efficiency.

Wyatt's 'missing digit' test forms an interesting modification of the computation test. In it examples in addition, subtraction, multiplication and division are given in which one or more figures, both in the answer and in the body of the example, are replaced by dots: the task is to restore the figures correctly. The following will serve to illustrate his material:

2.94
.867
781.
———
.42.6

REFERENCES

(1) H. Aikins, E. L. Thorndike and Elizabeth Hubbell, Correlations among perceptive and associative processes. *PsR*, 9: 1902, 374-382.

(2) Tsuru Arai, Mental fatigue. *ColumbiaConEd*, No. 54. New York, 1912. Pp. 115.

(3) Bellei, Ulteriore contributo allo studio della fatiga mentale nei fanciulli. *Revista sperim. di freniatria*, 30: May, 1904. See summary by Binet, *AnPs*, 11: (1904) 1905, 369.

(4) E. Bischoff, Untersuchungen über Uebungsfähigkeit und Ermüdbarkeit bei 'geistiger' und 'körperlicher' Arbeit. *ArGesPs*, 22: 1912, 423-452.

(5) W. Brown, Some experimental results in the correlation of mental abilities. *BrJPs*, 3: 1910, 296-322. (Forms 3d part of his Use of the theory of correlation in psychology. Cambridge, 1910. Pp. 83.)

(6) L. Burgerstein, Die Arbeitskurve einer Schulstunde. *ZSeGd*, 4: 1891, 543-564, 607-627. Also published separately in German, and in condensed form in English as: The working curve of an hour.

(7) L. Burgerstein und A. Netolitzky, Handbuch der Schulhygiene, 2d ed. Jena, 1902. Pp. 997.

(8) W. P. Burris, The correlations of the abilities involved in secondary school work. *ColumbiaConPhPs*, 11: 1903, 16-28.

(9) C. Burt, Experimental tests of higher mental processes and their relation to intelligence. *JEPd*, 1: 1911, 93-112.

(10) C. Burt and R. C. Moore, The mental differences between the sexes. *JEPd*, 1: 1912, 273-284, 355-388.

(11) W. G. Chambers, Individual differences in grammar-grade children. *JEdPs*, 1: 1910, 61-75.

(12) S. A. Courtis, The Courtis tests in arithmetic. *Final Rept. Educational Investigation Committee on School Inquiry, City of New York*. Vol. I, New York. 1911-1913. Pp. 397-546.

(13) M. E. Donovan and E. L. Thorndike, Improvement in a practise experiment under school conditions. *AmJPs*, 24: 1913, 426-8.

(14) H. Ebbinghaus, Ueber eine neue Methode zur Prüfung geistiger Fähigkeiten in ihre Anwendung bei Schulkindern. *ZPs*, 13: 1897, 401-457.

(15) W. S. Fox and E. L. Thorndike, The relationships between the different abilities involved in the study of arithmetic. Sex differences in arithmetical ability. *ColumbiaConPhPs*, 11: Feb., 1903, 32-40.

(16) J. Friedrich, Untersuchungen über die Einflusse der Arbeitsdauer und der Arbeitspausen auf die geistige Leistungsfähigkeit der Schulkinder. *ZPs*, 13: 1897, 1-53.

(17) W. H. Heck, A study of mental fatigue in relation to the daily school program. Lynchburg, Va., 1913. Pp. 28.

(18) W. H. Heck, A second study of mental fatigue in relation to the daily school program. *PsCl*, 7: 1913, 29-34.

(19) H. L. Hollingworth, The influence of caffein on mental and motor efficiency. *ArPs(e)*, No. 22 (*ColumbiaConPhPs*, 20, No. 4). New York, 1912. Pp. 166.

(20) H. L. Hollingworth, Correlation of abilities as affected by practise. *JEdPs*, 4: 1913, 405-414.

(20a) H. L. Hollingworth, Variations in efficiency during the working day. *PsR*, 21: 1914, 473-491.

(21) Marion E. Holmes, The fatigue of a school hour. *PdSe*, 3: 1895, 213-324.

(22) E. E. Jones, The influence of bodily posture on mental activities, N. Y., 1907. Pp. 60. (Reprinted from *ArPs(c)*, No. 6.)
(23) E. Katzen-Ellenbogen, The mental efficiency in epileptics. *Epilepsia*, vol. 3, 504-546.
(24) R. Keller, Ueber den 40-Minutenunterrichtsbetrieb des Gymnasiums u. der Industrieschule in Winterthur. *InMagScHyg*, 2: 1906, 298-330, especially 307-318.
(25) F. Kemsies, Arbeitshygiene der Schule auf Grund von Ermüdungsmessungen. *SmAbPdPs*, 2: 1898. Pp. 64.
(26) E. Kraepelin, Ueber geistige Arbeit, 2d ed. Jena, 1897. Pp. 29.
(27) E. Kraepelin, Die Arbeitscurve. *PhSd*, 19: 1902, 459-507.
(28) F. Krueger und C. Spearman, Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten. *ZPs*, 44: 1907, 50-114.
(29) H. Laser, Ueber geistige Ermüdung beim Schulunterrichte. *ZScGd*, 7: 1894, 2-22.
(30) H. D. Marsh, The diurnal course of efficiency (Columbia Univ. thesis), N. Y., 1906. Pp. 99.
(31) Gladys W. Martin, The evidence of mental fatigue during school hours. *JEPd*, 1: 1911, 39-45, 137-147.
(32) Gladys W. Martyn, A study of mental fatigue. *BrJPs*, 5: 1913, 427-446.
(33) E. Meumann, Vorlesungen zur Einführung in die exp. Pädagogik, *1st ed.*, 2 vols., Leipzig, 1907. Pp. 555 and 467.
(34) A. Oehrn, Exp. Studien zur Individualpsychologie. *PsArb*, 1: 1896, 92-151.
(35) M. Offner, Mental fatigue. (Eng. trans.) *EdPsMon*. Baltimore, 1911. Pp. 133.
(36) Louise E. Ordahl, Consciousness in relation to learning. *AmJPs*, 22: 1911, 158-213, especially 194-202.
(37) J. Reis, Ueber einfache psychologische Versuche an Gesunden u. Geisteskranken. *PsArb*, 2: 1899, 587-694.
(38) L. A. Robinson, Mental fatigue and school efficiency. *Bull. Winthrop Normal and Industrial College of South Carolina*, 5: 1911. Pp. 56.
(39) R. Schulze, 500,000 Rechenaufgaben. Eine exp. Untersuchung. *Praktische Schulmann*, 44: 1895, 340.
(40) R. Schulze, Experimental psychology and pedagogy. (Eng. trans.) New York, 1912. Pp. 364.
(41) B. R. Simpson, Correlations of mental abilities. *ColumbiaConEd*, No. 53. New York, 1912. Pp. 122.
(42) W. Specht, Ueber klinische Ermüdungsmessungen. *ArGesPs*, 3: 1904, 245-339.
(43) D. Starch, Transfer of training in arithmetical operations. *JEdPs*, 2: 1911, 306-310.
(44) Teljatnik, article in Russian. See for details, Burgerstein u. Netolitzky, 4; especially pp. 462-5.
(45) E. L. Thorndike, Mental fatigue. *PsR*, 7: 1900, 466-482, 547-579.
(46) E. L. Thorndike, Educational psychology, 2d ed., N. Y., 1910. Pp. 248.
(47) E. L. Thorndike, Measurements of twins. *Archives of Phil., Psych.* etc., No. 1, Sept., 1905. Pp. 64.
(48) E. L. Thorndike, The effect of practise in the case of a purely intellectual function. *AmJPs*, 19: 1908, 374-384.
(49) E. L. Thorndike, Practise in the case of addition. *AmJPs*, 21: 1910, 483-486.
(50) E. L. Thorndike, Mental fatigue. *JEdPs*, 2: 1911, 61-80.

(51) E. L. Thorndike, The effect of continuous exercise and of rest upon difficult mental multiplication. *JEdPs*, 5: 1914, 597-599.
(52) E. L. Thorndike, Measurements of ability to solve arithmetical problems. *PdSe*, 21: 1914, 495-503.
(53) E. L. Thorndike, Educational Psychology, Vol. III. (Mental work and fatigue, etc.) New York, 1914. Pp. 408, especially Chs. 2 and 3.
(54) R. Vogt, Ueber Ablenkbarkeit und Gewöhnungsfähigkeit. *PsArb*, 3: 1901, 62-201, especially 80-118, 131-135.
(55) F. L. Wells, Technical aspects of experimental psycho-pathology. *AmJIns*, 64: 1908, 477-512.
(56) F. L. Wells, The relation of practise to individual differences. *AmJPs*, 23: 1912, 75-88.
(57) F. L. Wells, Standard tests of arithmetical associations. *JPh*, 4: 1907, 510-512.
(58) W. H. Winch, Some measurements of mental fatigue in adolescent pupils in evening school. *JEdPs*, 1: 1910, 13-23, 83-100.
(59) W. H. Winch, Accuracy in school children. Does improvement in numerical accuracy 'transfer?' *JEdPs*, 1: 1910, 557-589.
(60) W. H. Winch, Further work on numerical accuracy in school children. Does improvement in numerical accuracy transfer? *JEdPs*, 2: 1911, 262-271.
(61) R. S. Woodworth and F. L. Wells, Association tests. *PsMon*, No. 57, 1911. Pp. 85, especially 44-48.
(62) S. Wyatt, The quantitative investigation of higher mental processes. *BrJPs*, 6: 1913, 109-133.

TEST 36

Mirror-drawing.—The preceding tests of association deal with *S*'s facility in producing unrestricted series, or in reproducing restricted series that have already been learned. The present test compels *S* to form a new series of associations that are opposed to associations stereotyped by several years of daily experience.

More particularly, in tracing an ordinary drawing the movements of the hand are guided by the visual perception of the drawing, plus kinesthetic sensations set up by the movement of the pencil. If the drawing is seen not directly, but in a mirror, the natural relations are reversed in certain respects, so that a new series of associative connections must be established between eye and hand. The rapidity and ease with which these new connections are established may be taken as an index of learning-capacity.

Learning is often said to take place either by practise (trial and error), by imitation, or by some form of ideational control (instruction, reasoning, etc.). In the mirror-drawing test, the conditions preclude the use of imitation, and there is but rela-

tively little opportunity to employ ideational control; whatever improvement appears is due primarily to a process of trial and error.

The interesting phenomena of mirror-writing are mentioned in psychological literature as early as the 90's, if not before, but the first use of mirror-drawing as a psychological experiment appears to be found in Henri's article on the muscular sense (9) and in his monograph on tactual space perception of the same year, 1898. W. F. Dearborn (7), independently, experimented with mirror-drawing in 1905, though his work was not reported until after other writers, likewise independently, had hit upon a similar idea. In addition to Dearborn, Judd (11, p. 99) Starch (16) and Hill (10) have called attention to the usefulness of mirror-drawing as a demonstration experiment to illustrate the acquisition of motor habits, the trial and error method of learning, the cross-transfer of practise-effects, and the like. Burt, Yoakum and Calfee, Miss Weidensall and others have used mirror-drawing to test quickness of learning,[1] and its correlation with sex, intelligence and other factors.

APPARATUS.—Mirror. Cardboard screen about 17x24 cm. Suitable supports for holding the cardboard. Thumb tacks. Stop-watch. Two kinds of diagrams, printed in red ink, for tracing: (*a*) a 6-pointed star, (*b*) a set of 6 patterns, each based upon a group of 12 points arranged at equidistant intervals in a circle about its central point, with guiding lines joining the 13 points in irregular fashion. [Mechanical counter. A strong prism (about 20 D.).]

The six patterns resemble those used by Miss Calfee and Yoakum, but differ from them in eliminating the numbering of the points. *S* is directed from one point to the next by means of arrows and broken red lines, and does not have, therefore, to spend a portion of his time in hunting for the numbering to discern the order in which the points are connected. These patterns are somewhat more difficult than the star in so far as the direction of the movement of the hand is not the same in the various trials—a matter of advantage, however, since the series of trials does not develop a specific memory for a given set of hand movements. On the other hand, these patterns are less difficult than the star in so far as *S* is not required to follow the directing lines exactly. The patterns also more nearly resemble than does the star test the mirror test used by Burt, in which *S*

[1]For a discussion of prolonged adaptation to mirror-vision, see Stratton (18).

was required to punch with a stylus through 8 or more orifices arranged in a circle about an orifice at the center. They are superior to Burt's material in several respects.

Whether the stars or the patterns are used will depend upon circumstances. Either form of material may be used as supplementary to the other; thus, the star test may be used before and after drill work with the 6 patterns for an experiment to test the effect of practise.

PRELIMINARIES.—Pin the diagram out flat upon a table, directly in front of *S*. If the patterns are used, they should be taken in the order of their numbering. If the star is used, it should be placed with the cross-line that indicates the starting-point at the back (away from *S*) and with the card square with the edge of the table. (This brings the star slightly 'out of true,' as is intended.) Set up the mirror inclined slightly (about 5 deg.) from the vertical, just beyond the diagram. Arrange the screen (see Fig. 64) so that it will cut off *S*'s direct view of the diagram, but will allow him to see it clearly in the mirror, and will not interfere with his hand in drawing.

METHOD.—(*a*) *With the patterns.* Place the point of a lead pencil at the center of the diagram. Assist *S* to grasp the pencil (permitting him to look only in the mirror). Instruct him: "When I say 'now,' move your pencil along the paper in the direction indicated by the red arrow till you reach the point at the end of the broken line; then follow the red line from that point to the next one, and so on till you have touched all 12 of the points on the paper and come to the end of the red dashes. You don't have to keep on the lines; they are put there simply to show you where to look for the points, but you must keep your pencil on the paper, and you must bring your pencil to each point before you go on to the next one. Work as rapidly as you can. Don't stop to figure out what you ought to do, but keep your pencil moving all the time." Start the watch at the signal, and record the time for the entire diagram. Pin down the second pattern and continue until all six patterns have been traced.

(*b*) *With the star.* Place the point of a lead pencil upon the cross-line of the star, and assist *S* to grasp the pencil (permitting him to look only in the mirror). Instruct *S:* "Trace the outline of the star, starting in this direction [indicating, *by pointing,* the tip of the star at the right of the cross-line].

FIG. 64. THE MIRROR-DRAWING TEST.

Work as rapidly as you can, but *try to keep on the line.* Don't stop to figure out what you ought to do, but keep your pencil going in some direction, and keep its point on the paper all the time." Start the watch, and record the time for the entire drawing.

E may also note the time for each sixth of the pattern. But it is, perhaps, more desirable to supplement the total time by a record of the total number of corrective movements made by *S*. Since these movements are often rapid, and of short extent, it is necessary to use a mechanical or other form of counter to obtain the record. Press the counter every time *S* moves toward the line.[2]

For a standard test, make 6 trials with the right hand, using a fresh star for each trial.

VARIATIONS OF METHOD.—Make tests with the star before and after a drill series with the patterns, as suggested above, or

[2]Every 'error,' or movement away from the line must, of course, be compensated for by a return movement. The idea is to register the number of these errors, or corrective movements. Changes of direction necessitated by the pattern itself are, obviously, to be neglected.

with either form of material make a first trial with the left hand; follow with a series of 5 to 50 trials with the right hand,[3] then return to the left hand for a final test. Note how much practise effect has been 'transferred' from the one hand to the other. Plot a graph to show the effect of practise, both upon the time and upon the corrective movements.

TREATMENT OF DATA.—In the standard form of test *E* has available 6 records. Several possibilities appear: *S*'s may be compared with respect to (1) their 1st trial, (2) their 6th trial, (3) all 6 trials taken collectively (sum or average), or with respect to their rate of improvement, by computing the per cent. of gain either (4) in the 6th, compared with the 1st trial, or (5) in the average of the last three, compared with the average of the first three trials. The third method was found by Burt to yield the best correlation with intelligence. On the other hand, the 2d method would seem to have some merit, since Yoakum and Calfee conclude that "the time consumed in the first trial is an individual variation; that of the last [6th] more nearly represents the individual's place in the group." Until we have more investigations on this point it would be better for *E* to try more than one method of ranking *S*'s and to select the one which gave the most favorable results.

RESULTS.—(1) The best *norms* for the patterns are supplied by the results of Yoakum and Calfee, embodied in Table 78; results from a more limited number of college students with the star test are shown in Table 77, and for other groups in Tables 79 and 80.

TABLE 77

Effect of Practise on Speed in Mirror-Drawing. College Students (Whipple)

	NUMBER	1ST LEFT	1ST RIGHT	2D RIGHT	3D RIGHT	4TH RIGHT	5TH RIGHT	2D LEFT
Men	11	169	127	108	96	80	67	88
Women	23	149	127	87	76	67	67	74

[3]If desired, the 6 patterns may be used, turned to bring the other edges at the back, in order to provide drill without direct repetition of the same diagram.

(2) *Individual differences* in performance are striking; thus, in the star test the time consumed in making the first tracing ranged, in the author's tests of 34 students, from about 50 sec. to more than 8 min. In the larger group of students examined

TABLE 78

Times, in Sec., for Mirror-Drawing (Yoakum and Calfee)

GROUP	TRIAL	I	II	III	IV	V	VI	AVER.
I----	Median	243.0	121.0	93.0	82.0	68.0	50.0	110.33
	M. V.	94.9	45.5	28.1	34.7	24.7	17.1	36.57
	Slowest	517.0	245.0	205.0	180.0	158.0	113.0	210.00
	Fastest	69.0	51.0	41.0	43.0	40.0	32.0	53.66
II----	Median	92.0	65.0	48.0	41.0	35.0	28.0	54.70
	M. V.	64.1	33.9	26.6	19.3	21.9	14.2	27.40
	Slowest	700.5	337.5	303.5	153.5	201.8	171.0	242.37
	Fastest	31.5	23.5	19.3	18.3	17.8	17.0	23.95
III----	Median	167.5	105.0	80.0	68.0	56.0	48.0	97.83
	M. V.	104.2	39.3	30.3	19.7	19.9	13.5	33.38
	Slowest	752.0	277.0	270.0	175.0	121.0	105.0	193.33
	Fastest	72.0	49.0	40.0	34.0	33.0	23.0	46.87

Group I comprised 30 elementary school boys. Group II, 52 women, and Group III, 51 men in the freshman class of the University of Texas. The averages for each group in each trial are not here reproduced.

at Texas differences range from 31.5 to 752 sec., while the fastest college girl tested by Miss Weidensall had a record of 18 sec., as compared with 2072 sec. for the slowest reformatory woman (Table 79). These differences, as inspection of the tables will show, are greatly reduced after a little practise.

(3) *Dependence on sex.* That girls decidedly surpass boys and that women decidedly surpass men is shown in all the published results in mirror-drawing, with the exception of two groups reported by Burt and Moore, and in them certain divergencies in method and in other test conditions offer a sufficient explanation of the apparent exception. Miss Calfee's averages for six trials give for the freshmen women 64.4 sec., P.E. 22.3, for the freshmen men 101 sec., P.E. 28.5. She finds that only 6 per cent. of the men reach the women's median, while 90.4

per cent. of the women reach the men's median. It is not only possible, but probable, that this sex-difference is in some part due to greater familiarity of women with the use of the mirror. Burt believes that there is also an innate sex difference at work.

(4) *Dependence on practise.* (*a*) *General practise-effects.* The tables given above show that even a single trial produces a decided reduction in time: the median time for elementary school boys, for example, is cut in halves in the pattern test, while that for men and women is reduced one-third by the first trial. (See Fig. 65.) The long practise experiment conducted by Starch with the star test shows (Fig. 66) that the reduction is rapid at first, then slower, and that maximal speed is not attained for a long time, apparently not until some 90 trials (Starch's curve represents a series of 100 trials, one per day).

(*b*) *Individual differences in practise-effects.* Practise curves compounded of the performances of a group of *S*'s show a smooth drop (see Fig. 65), but the curves of individual *S*'s are not necessarily of this form: on the contrary, it is possible, as Yoakum and Calfee have shown (22, p. 290), to separate *S*'s into groups that show the 2d trial slower than the 1st, or the 3d slower than the 2d, etc. These investigators summarize these facts by saying: "Some *S*'s gain control of the situation by a fairly regular procedure; others temporarily lose control at some point in the series. The majority of the latter lose control at the fourth or fifth trial in a series of six tests." It follows that the rank-order of *S*'s in any one trial does not correlate perfectly with their rank-order in any other trial; actual correlations computed by Yoakum and Calfee between the first and subsequent trials are 0.79, 0.76, 0.74, 0.64 and 0.59 for the 2d, 3d, 4th, 5th and 6th trial, respectively. The correlation between the first trial and the average of all 6 trials is given by them as 0.93.

(*c*) *Cross-education.* A considerable amount of practise gained with the one hand is transferred to the other (unpractised) hand. Thus, Starch's 100-day practise with the right hand effected an improvement in it of 92 per cent. in accuracy and of 84 per cent. in speed. A single left-hand record, made

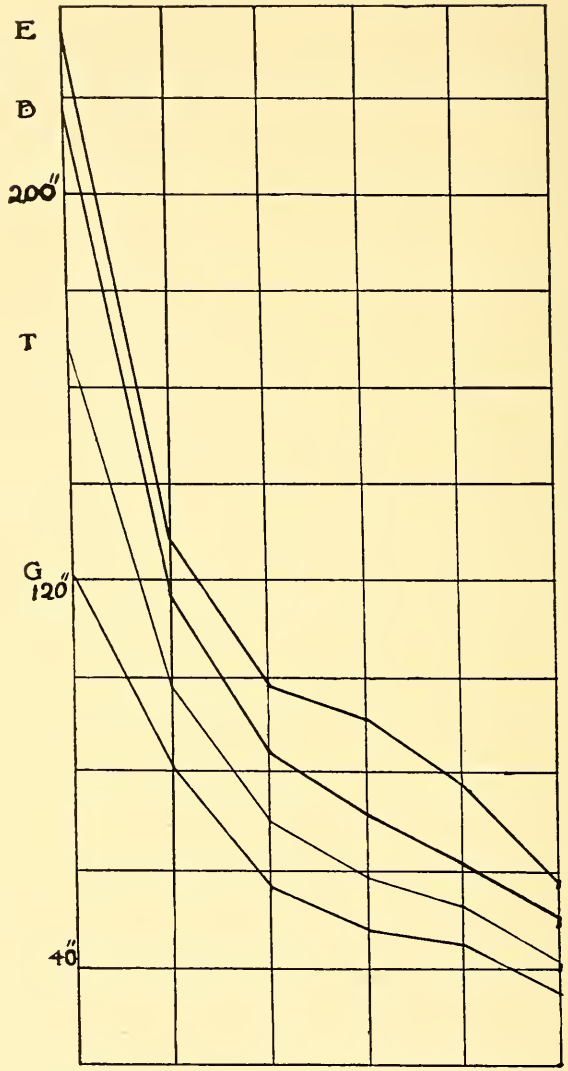

FIG. 65. THE EFFECT OF PRACTISE DURING SIX TRIALS UPON THE AVERAGE SPEED OF MIRROR-DRAWING. (From Yoakum and Calfee)

The abscissas represent the six trials, running from left to right. The ordinates represent time in sec. 'E' is the curve for the 30 elementary school boys, 'B' for the 51 freshmen men, 'G' for the 52 freshmen women, 'T' for the 103 freshmen collectively.

at the expiration of this period, showed, in comparison with a single left-hand record made before practise began, an improvement of 81 per cent. in accuracy and of 85 per cent. in speed. There is, however, nothing surprising in this so-called 'cross-

FIG. 66. THE EFFECT OF LONG PRACTISE UPON MIRROR-DRAWING.
(From Starch)

education,' since the tracing of the star in the mirror depends primarily upon co-ordinations established in the central nervous system: in other words, the transfer is only an outwardly apparent transfer; in reality, the same factors are at work in the control of either hand.

(*d*) *Persistence of practise.* The effect of even a short period of practise in mirror-drawing is very persistent. Thus, Burt administered 6 tests in succession, during which the average speed fell from 103 to 39.5 sec. Twelve weeks later, two tests were given in succession; the average speed developed was 34.5 sec. in the first, and 27.4 sec. in the second: in other words, the 7th test surpassed the 6th, made 12 weeks previously—a condition found in the records of 16 out of 26 boys. The extent to which this persistence of practise-effect was shared by Burt's

S's is further indicated by the correlation of 0.52 between their standing before, and their standing after the 12-week interval.

Hill's work (10b) shows that the skill developed by one trial a day, continued for 50 days, is so persistent that after an interruption of three years the first trial in relearning is as fast as the 32d and more accurate than the 50th trial of the original series, and that in four retrials a speed and accuracy has been regained that is equal to the final records of the original series. Mirror-drawing seems, therefore, to resemble neuro-muscular habits, like skating, typewriting, etc., in the manner in which skill once developed is retained with little loss over long periods, rather than the associative connections of ideational life with their relatively lesser persistence.

(5) *Dependence on intelligence.* Burt reports a correlation between speed and estimated intelligence of 0.67, P.E. .07, for elementary school boys, and of 0.54, P.E. .14, for preparatory school boys. In another group of English school children a correlation of 0.60 was found, according to Burt and Moore. Miss Calfee, however, found no such relations in her group of elementary school children chosen to duplicate Burt's conditions: here the correlation with school grades was virtually zero (0.07); similarly, in the college students the correlation with grades was —.07 in the case of the men and 0.19 in the case of the women. The author was able to discern no constant differences between the work of five dull and five bright boys.

(6) *Delinquents.* Comparative study of the star-test (5 successive trials) with college girls, maids in college dormitories and girls at Bedford Hills, N. Y., Reformatory, conducted by Miss Weidensall, reveals a number of interesting results. From advance sheets of her manuscript, for which I am indebted to Dr. Weidensall, I have selected data referring primarily to the time records only[4] (Tables 79 and 80). The first of these tables shows that, both in the first and in the last trial, and whether

[4]Dr. Weidensall expects to publish also data for the number of errors (corrective movements) and for the degree of 'precision' with which the line is followed. Precision has been measured by ascertaining the total number of cm. in the contour of the star in which, for distances of 3 or more consecutive mm., the tracing remained within 2 mm. either way from the red line.

maximal, minimal, median, average or upper or lower quartile is considered, the three groups are invariably arranged in the same order—students best, Bedford women last and the maids intermediate. The second of these tables shows that there

TABLE 79

Times, in Sec., Used in the Star Test by 36 College Girls, 16 College Maids and 69 Bedford Reformatory Women (Weidensall)

	COLLEGE GIRLS		COLLEGE MAIDS		REFORMATORY WOMEN	
	1st Star	5th Star	1st Star	5th Star	1st Star	5th Star
Fastest	18.	7.	36.	21.	59.	36.
Upper Q.	41.	17.5	54.	29.5	203.6	80.
Median	66.	28.7	127.5	44.5	420.	117.2
Average	82.6	31.3	133.6	48.6	473.1	124.
Lower Q.	110.	39.	161.	69.	627.	148.
Slowest	252.	76.	409.	85.	2072.	436.*

*With two failures in addition.

exists a good correspondence between both the time and errors for the star test and the classification made by the institution into three groups depending on outlook for reformation: the differences are more striking in the first than in the fifth tracing.

TABLE 80

Scores in the Star Test for Three Groups of Bedford Reformatory Women (Weidensall)

INSTITUTIONAL CLASSIFICATION	FIRST STAR		FIFTH STAR	
	Time	Errors	Time	Errors
Most capable and promising	320.9	117.7	105.4	36.0
Women with illegitimate children under 2 yrs. of age	562.9	211.3	123.1	45.6
Backward and mentally feeble. Unpromising	610.5	264.4	127.2	55.1

In addition to these quantitative results, the star test has proved to possess a value in a perhaps unexpected direction,

viz.: as a device for sorting out S's of the unstable and less tractable type.

On this point Dr. Weidensall writes: "This test isolates better than any we have tried at Bedford those who are incapable of sustained effort under difficulties. It isolated, of course, the low-grade feeble-minded, for, no matter how hard they try, they do not succeed in tracing a precise star. The epileptics have a characteristically bad time and their stars are all 'knotted up' with 'blind spots' where they were caught and held indefinitely. Chiefly, however, is the test of interest in the case of those who are bright enough to trace the star well, but too unstable to do so. These are invariably the girls who are difficult to manage in the institution. The tracing goes well enough until suddenly the pencil at some hard point starts off in the wrong direction. The subject then tugs and pulls, grows more and more irritated, disturbed and excited, makes big black circles and finally throws down the pencil and gives up. When calmed, praised and urged to try again, she will continue and usually in the end draw a fairly good fifth star. This behavior in tracing the star is typical of their behavior in the institution when the pressure of discipline or responsibility becomes the least bit too exacting."

(7) *Relation of speed and accuracy.* The curves reproduced from Starch show that practise produces a reduction in the number of corrective movements that parallels fairly closely the reduction in time. Correlations between time and errors obtained by Miss Weidensall are for the students 0.63, for the maids 0.87, for the reformatory women 0.61. My own work with college students has given a correlation of 0.86, P.E. .04.

(8) *Reliability.* Burt and Moore give this coefficient as 0.52. The method used at Texas is evidently superior, since the coefficient of relation between the first and second test, as above stated, amounts to 0.79, and thus assures satisfactory reliability.

(9) *Various correlations.* Miss Calfee's tests of Texas freshmen included three tests previously used by Burt, viz.: card dealing, card sorting and alphabet sorting. Correlations found by Burt between mirror-drawing and these three tests when applied to school children were 0.40, 0.34 and 0.29, respectively: those found by Miss Calfee for school children were only 0.11, 0.26 and 0.06, for freshmen men 0.19, 0.11 and 0.22, and for freshmen women 0.37, 0.20 and 0.29, respectively. Save, then, for the last mentioned correlation, her figures are invariably lower than those of Burt. Other 'corrected' correlations re

ported by Burt for mirror-drawing (average correlations for various groups) are: tapping 0.74, dotting apparatus 0.92, spot-pattern test 0.75, immediate memory 0.38, discrimination of pitch 0.66, comparison of line lengths 0.55, esthesiometer 0.38, discrimination of lifted weights 0.30.

(10) *Qualitative aspects.* Efficiency in mirror-drawing may result from the actual formation of new visual-motor co-ordinations (indeed, some *S*'s after executing a number of drawings, find that, for a short time immediately thereafter, these new co-ordinations interfere with normal drawing or writing); but efficiency may also result, at least in the star test, from the voluntary inhibition of visual control in favor of kinesthetic control, *i. e.*, by thinking the drawing of a star in motor terms, as if working with the eyes shut. Or, the hand-movements may be started in this manner and then carried out by visual control from the mirror. Finally, adults occasionally control the drawing ideationally, *i. e.*, by applying inferred properties of reflection by mirrors.

It is evident that the existence of these qualitative differences may affect the test in such a way that the quantitative data for different *S*'s may 'measure' different mental processes.

Very slow *S*'s get 'caught' at certain difficult points of the drawing, where they make a long series of futile attempts to start in the right direction. Here the normal visual-motor control is too persistent to be readily broken or ignored.

NOTES.—A further study of the associative connections involved in mirror-drawing may be made by the use of dot-tapping through a prism or of the various forms of mirror-writing.[5]

For the first test, let *S* shut his left eye, and strike repeatedly with his right forefinger at a mark on the wall or table-top,

[5]On mirror-writing, consult Abt, Allen, Downey, Laprade, Lochte, Ordahl, Rowe, Strack, Weber, and Wegener. The most elaborate statistical study is that of Lochte, who examined 2804 pupils in Berlin, and found, for children aged 6-7 years, 13.2 per cent. of spontaneous left-hand mirror-writing in boys and 25.4 per cent. in girls, but for children aged 13-14 years, only 0.7 per cent. in boys and 35 per cent. in girls. The tendency toward this type of writing appears, therefore, to decrease with age, and to be more evident in girls than in boys.

The most elaborate qualitative analysis of the various 'controls' used in writing is that of Miss Downey.

making about one stroke per second, after the manner prescribed in the test of precision of aiming (No. 11). After this rhythmic movement has become well established, and *without interrupting it in the least*, place suddenly before his eye a 20 D. prism, with the base toward his nose. The mark is thereby apparently displaced some 10 cm. to the left. Count the number of strokes that *S* makes before he hits the mark again (with the prism kept before the eye). Similarly, count the number of strokes necessary to hit the mark again when the prism is removed.

For the second test, try any or all of the following:

(1) Close the eyes and write with both hands simultaneously. Cases will then appear, particularly in young children, of spontaneous mirror-writing (writing which reads correctly when held before a mirror) with the left hand. If this appears, see if *S* can write normally with the left hand when his eyes are closed.

(2) Show *S* a sample of mirror-writing. Explain its nature. Ask him to write in a similar manner, first with his left, then with his right hand.

(3) Write with both hands simultaneously, but with the left intentionally in mirror-writing.

(4) Read normal writing when seen only as reflected in a mirror.

(5) Write normally while watching the writing in the mirror, *i. e.*, with hand and paper hidden from direct observation, as in the star test.

REFERENCES

(1) G. Abt, L'écriture en miroir. *AnPs*, 8: 1901 (1902), 221-225.

(2) F. J. Allen, Mirror-writing. *Brain*, 19: 1896, 385-7.

(3) C. Burt, Experimental tests of general intelligence. *BrJPs*, 3: December, 1909, 94-177, especially 145-9.

(4) C. Burt, Experimental tests of higher mental processes and their relation to general intelligence. *JEPd*, 1: 1911, 93-112.

(5) C. Burt and R. C. Moore, The mental differences between the sexes. *JEPd*, 1: 1912, 273-284, 355-388.

(6) Marguerite Calfee, College freshmen and four general intelligence tests. *JEdPs*, 4: 1913, 223-231.

(7) W. F. Dearborn, Experiments in learning. *JEdPs*, 1: 1910, 373-388.

(8) June E. Downey. (*a*) Control processes in modified hand-writ-

ing: an experimental study. *PsMon*, 9: April, 1908, No. 37. Pp. 158. (*b*) On the reading and writing of mirror-script. *PsR*, 21: 1914, 408-441.

(9) V. Henri, Revue générale sur le sens musculaire. *AnPs*, 5: 1898 (1899), 399-513, especially 504-508. (Also Ueber die Raumwahrnehmungen des Tastsinnes. Berlin, 1898. Pp. 228, especially p. 140.)

(10) D. S. Hill (*a*) Class and practise experiments upon the learning process. *PsB*, 8: 1911, 70-71. (*b*) Minor studies in learning and relearning. *JEdPs*, 5: 1914, 375-386.

(11) C. H. Judd, Laboratory manual of psychology. New York, 1907.

(12) A. Laprade, Contributions a l'étude de l'écriture en miroir. 1902.

(13) Lochte, Beitrag zur Kenntnis des Vorkommens u. der Bedeutung der Spiegelschrift. *Arch. f. Psychiatrie u. Nervenkrankheiten*, 28: 1896, 379-410.

(14) Louise E. Ordahl, Consciousness in relation to learning. *AmJPs*, 22: 1911, 158-213, especially 193f.

(15) E. C. Rowe, Voluntary movement. *AmJPs*, 21: 1910, 513-562, especially 537ff.

(16) D. Starch, A demonstration of the trial and error method of learning. *PsB*, 7: January, 1910, 20-23.

(17) M. Strack, Mirror writing and left-handedness. *PdSe*, 2: 1893, 236-244.

(18) G. M. Stratton, The spatial harmony of touch and sight. *Mind*, n. s. 7: 1899, 492-505.

(19) H. Weber, Spiegelschrift u. Lenkschrift. *Zeits. f. klin. Med.*, 27.

(20) H. Wegener, Die Spiegelschrift. *ZPs*, 1: 1899, 254-269.

(21) Jean Weidensall, The mentality of the criminal woman. To appear in *EdPsMon*.

(22) C. S. Yoakum and Marguerite Calfee, An analysis of the mirror-drawing experiment. *JEdPs*, 4: 1913, 283-292.

TEST 37

Substitution.—This test is one of many that may be devised to measure the rapidity with which new associations are formed by repetition. The name commonly applied to the test arises from the process that it involves, in which *S* is called upon to substitute for one set of characters (letters, digits, familiar geometrical forms, etc.) another set of characters in accordance with a plan set before him in a printed key. The procedure differs from most memory tests or exercises of memorizing in that the connections indicated by the key are not committed to memory at the outset, but acquired gradually by use as the test proceeds.

The principle embodied in such a test obviously admits of numerous variations in detail of application. One form of substitution, the replacement of a set of letters by another set of letters, was used by Lough (7) in 1902 for a class exercise in learning. Another and more elaborate form in which letters

distributed like those of a typewriter keyboard are to be associated to numerals is reported by Starch and Dearborn to have been devised by Jastrow and used several years ago in the Wisconsin University Laboratory. In recent years several variations, some simpler, some more difficult, have appeared.

The substitution test seems primarily to have been developed as a useful demonstration and class experiment for the study of the psychology of learning and of the practise curve (Dearborn, Starch, Lough, Munn, Kline). It has also been used to study racial differences (Baldwin, Pyle), to trace the effect of dental treatment on general ability (Kohnky), to compare delinquent and normal individuals (Baldwin, Weidensall) and as one test of the capacity of working children (Woolley and Fischer). Incidentally, of course, the relation of learning ability to age, sex and school training has been the object of investigation.

Three forms of test material are here presented: the first and second, which are modifications of a form devised by W. F. Dearborn (3), may be used with adults or older children; the third, which has been devised by Mrs. Woolley and used by Miss Kohnky and Miss Weidensall as well as Mrs. Woolley, is much simpler and better adapted for younger or less capable children.[1]

A. STANDARD FORM FOR INDIVIDUAL PROCEDURE (SYMBOL-DIGIT TEST)

MATERIALS.—Stop-watch, preferably split-second. Cover-board with key. Test strips.

The cover-board, about 18 x 36 cm., is so constructed as to furnish a sort of 'tunnel' through which the test-strips may slide as fast as they are written: it also carries a printed key consisting of 9 circles, within each one of which is a digit (from 1 to 9) and a symbol (square, asterisk, etc.)

The test-strips, about 11.5 x 50 cm., contain forty 5-place series of symbols like those of the key, together with forty 5-place empty squares.

[1] If a test more difficult than any of those described here is desired, reference may be had to the form proposed by Gray (4) and used with some modification by Baldwin (1). The Maltese Cross test arranged by Mrs. Squire, and also tried by Carpenter, proved undesirable, apparently in part because it was too easy. Much the same thing may be said of her Colored Forms test, which represents still another quite simple variety of substitution test.

METHOD.—Lay the cover-board upon the table. Insert a test-strip in such a manner that the first (top) line of characters comes just below the lower edge of the cover and hence just beneath the key.

Cover the key and do not allow *S* to examine it before the test, save as specified below.

Give *S* the following explanation: "You will find before you on the table a card on which there are nine circles. In each circle you will find one of the numbers from 1 to 9, and a symbol, *i. e.*, a small character or drawing. Then, you will find a strip of paper with rows of the same characters, and with empty squares beside them. What you are to do is to write in these empty squares the numbers that correspond with the characters. Keep at work continuously, as fast as you can, until you have filled in all the empty squares on the paper. Of course, you will have to look back and forth from the paper to the circles to find out what number to use, unless you can, after a while, remember some of the numbers without looking at them."

With young *S*'s, this verbal explanation will be insufficient to make the task clear. It will do no harm, in such cases, to show *S*, for a brief instant, the card of circles and a test-strip that has already been filled out. Let him see them just long enough to make the instructions clear, but not long enough to permit him to learn any of the combinations.

Start the watch when *S* starts the first line: keep the watch in view, but out of *S*'s sight: record, without stopping the watch, the position of the second-hand when *S* completes every 5th line (indicated, for this purpose, by a heavier division-line in the test-strip).

As fast as *S* finishes a line (or two lines), push the strip forward to bring a fresh line of symbols into position at the lower edge of the cover.

When the 40th line is written, conceal the key; immediately turn over the test-strip, write on it the digits 1 to 9, and ask *S* to place above each digit the character that accompanies it. Ascertain, if possible, whether *S* relied upon visual, auditory, visual-auditory, or some other type of associative imagery.

TREATMENT OF DATA.—Check up the test-strip for errors. Compare *S*'s with respect to (1) their time for the whole test, (2) their gain in the last, as related to their speed in the first 5-line section, (3) their accuracy, and (4) their knowledge of the symbols (crediting 1 for each symbol correctly reproduced, and 1 for each pair of transposed symbols). Plot graphs showing the variation in speed for the eight sections.

B. FORM FOR GROUP TESTS, OR FOR SUPPLEMENTARY INDIVIDUAL TESTS (DIGIT-SYMBOL TEST)

MATERIALS.—Printed form, at the top of which are shown 9 circles, as in Form A (save that different symbols are used), and in the body of which is provided, in two columns, a series of forty 5-place numbers and forty 5-place blank squares in which the appropriate symbols are to be placed. Stop-watch.

METHOD.—For individual tests, give instructions similar to those for Form A, with such modifications as the altered arrangement of the material necessitates. Make clear, especially, that the second column is to be filled out the moment that the first is completed.

For group tests, supplement the instructions by an adequate blackboard explanation, preferably with an illustration so devised as not to give information concerning the symbols to be used. Have the papers distributed, face down, to be turned over only at the command to start. Work by the time-limit method, allowing 4 min. for the test. Instruct *S*'s to place an oblique mark at the point reached when the command "mark" is heard. Give this signal every 30 sec., so that the work is divided into 8 periods of 30 sec. each. Conclude with the symbol-test as in the individual method. Plot curves for 30 sec. intervals.

VARIATIONS OF METHOD.—(1) Cut off the top of the form and glue the pattern of circles on a sheet of cardboard, as in Form A. Cut and paste the two test-columns to form a single long column, as in Form A. This will permit check-tests, comparable with the standard method, save that here symbols, there digits are written.

(2) Repeat either Form A or Form B after an interval of several hours, days, or weeks, to compare the permanence, in different *S*'s, of the associative connections established in a single trial.

(3) Repeat Form A until the associations are firmly established, and the digits can be written rapidly without seeing the pattern. Ascertain whether the use of Form B will then develop interference of associations.

(4) Cover up the key in either Form A or Form B when the last section (last quarter or last eighth) of the test is reached so as to produce a test of *S*'s ability to continue the work from memory, like that described for Form C.

C. CINCINNATI SYMBOL-DIGIT TEST.

MATERIALS.—Four test sheets of geometrical forms, each containing ten rows, 5 units per row, of nine different forms. Cardboard with printed key. Cardboard cover. Stop-watch.

METHOD.—Put before *S* the first test sheet and set the key where it can be seen easily. The following are the instructions then given by Woolley and Fischer:

"You see this page of figures [forms]. Now on this card I have the same figures, but each figure has a number in it. What I want you to do is to write in each figure on this page the number that you see in the same figure on that card. For instance, what figure would you put in here? [*E* points to one of the figures which might easily be confused with another one—the inverted triangle or the U, and corrects *S* if he makes a mistake.] And in here [pointing to one of the 'unique' figures]? I want you to begin here at the top of the page and fill the figures in, in rows, just as you come to them. As you finish each row, I will cover it up with this piece of cardboard, this way. Now begin, and see how fast you can get the whole page done."

The time is taken from the moment *S* begins to look on the key for his first number to the moment he writes the last one. The second test sheet is then given with the instruction: "Now fill in this page the same way, and see if you can do it faster this time."

The third test sheet follows, with the instruction: "Fill in this page and try to do it still faster. When you finish this page, I will take the card away, and then I want you to try to fill in the last page just from memory."

S is allowed to correct any errors that he may note before the line is covered. The covering is done to insure that each line in the first three sheets is done from the key and each line in the last sheet from memory, never from the previous records.

VARIATION OF METHOD.—If *S* scores less than 98 per cent. accuracy on Sheet 4, it is instructive to give another drill sheet, followed by a second test of substitution from memory, and to continue alternating sheets filled in with the key and without the key until this degree of accuracy is secured. The number of extra trials needed forms a useful indication of relative learning capacity, especially in the case of rather incompetent *S*'s.

TREATMENT OF DATA.—For each test sheet, taken separately, is figured the time, the accuracy and an index of efficiency computed from the time and the accuracy. Accuracy is calculated by subtracting from 100 per cent. 2 per cent. for each error or omission. The index is found by dividing the obtained time by the accuracy. In the first three sheets this index may be regarded as indicating approximately the time needed to make the substitutions without error. In the fourth sheet the index is evidently a more arbitrary measure, since an error in substituting from memory might not be remedied by any amount of extension of the time.

Speaking generally, the learning capacity of a given *S* is indicated not alone by his performance with the 4th sheet, but also by his index for the first three sheets, *i. e.*, while the 4th sheet shows whether the associative connections have been made correctly or not, the work with the other sheets shows how long a time was used in establishing these connections.[2]

RESULTS.—(1) *Norms* for the three substitution tests are now available in sufficiently satisfactory form for most purposes.

[2] It would seem possible that some measure of learning capacity might be calculated from the relation between performance with the 4th and with the other sheets, though the Cincinnati investigators have contented themselves with the treatment quoted.

Tables 81 and 82 give results for college students with Form B. Tables 83 and 84 give Pyle's results with Form B and Form A, respectively, for both sexes and ages from 8 years upward. Table 85 gives some of the more important norms compiled at Cincinnati for 753 children 14, and 679 children 15 years old

TABLE 81

Substitution Test. Number of Symbols Written. Form B. Group Method (Whipple)

THIRTY-SEC. PERIOD	1ST	2D	3D	4TH	5TH	6TH	7TH	8TH	TOTAL	SYMBOL SCORE
Average, 12 men	13.7	16.1	14.6	16.3	14.8	17.2	16.7	17.9	127.3	8
Average, 28 women	13.9	15.4	16.0	17.9	16.0	17.0	16.8	19.0	132.0	8.2
Fastest individual	10.0	21.0	22.0	18.0	23.0	20.0	25.0	26.0	165.0	9
Slowest individual	11.0	13.0	8.0	10.0	12.0	10.0	11.0	13.0	95.0	3

TABLE 82

Substitution Test. Speed in Seconds. Form B. Individual Method (Whipple)

SECTION OF 5 LINES	1ST	2D	3D	4TH	5TH	6TH	7TH	8TH	TOTAL
Average, 13 men	54.0	46.0	45.8	44.8	46.1	44.4	47.7	44.3	373.1
Average, 5 women	45.8	41.2	40.6	38.6	43.4	37.6	36.6	35.0	318.8
Total, 18 cases	51.8	44.7	44.3	43.1	45.4	42.5	44.6	41.7	358.1
Fastest individual	42.0	35.0	33.0	30.0	36.0	29.0	31.0	34.0	270.0
Slowest individual	63.0	58.0	59.0	61.0	62.0	53.0	60.0	65.0	481.0

TABLE 83

Correct Substitutions Made in 60 Sec. Digit-Symbol Test (Pyle).

SEX	AGE	8	9	10	11	12	13	14	15	16	17	18	ADULT
Male	Cases	34	58	50	49	56	62	48	35	31	14	17	67
	Aver.	10.3	12.6	15.4	16.3	19.1	22.6	21.1	24.7	24.8	23.8	28.7	29.3
	A. D.	3.5	4.1	3.9	3.6	5.1	5.8	4.5	4.6	5.4	4.3	3.5	8.7
Female	Cases	37	61	58	49	68	49	46	34	46	38	29	88
	Aver.	13.0	15.7	18.8	18.5	22.7	23.4	26.8	26.8	27.5	28.5	25.9	32.2
	A. D.	3.2	4.1	4.4	4.1	4.9	5.2	5.0	4.7	5.3	5.7	7.0	4.2

TABLE 84

Correct Substitutions Made in 60 Sec. Symbol-Digit Test (Pyle).

SEX	AGE	8	9	10	11	12	13	14	15	16	17	18	ADULT
Male	Cases	37	72	76	62	75	78	59	45	38	20	17	56
	Aver.	10.0	13.2	16.5	17.7	19.3	20.7	23.3	25.8	27.8	26.1	28.0	33.0
	A. D.	5.3	5.0	5.8	5.4	5.4	5.7	5.4	5.9	6.3	7.4	5.1	9.3
Female	Cases	41	82	82	63	89	66	62	44	55	43	29	89
	Aver.	10.9	16.0	19.9	19.6	23.1	25.6	27.4	29.7	29.1	32.0	33.1	31.3
	A. D.	5.3	5.2	6.4	6.3	6.6	6.4	6.1	6.7	5.3	6.3	4.4	5.4

applying for working certificates.[3] From data kindly supplied me by Mrs. Woolley I have constructed also the percentile curves, Figs. 67 and 68, for the same groups of children.

(2) *Dependence on age.* Pyle's averages, with two exceptions, show that the capacity in the substitution test improves every year from 8 to 18, both in boys and in girls. The Cincinnati children at 15 surpass their 14-year-old records, with every page and in both speed and accuracy: the difference is too pronounced to be due to the repetition of the test, since different keys were employed in the two trials.

(3) *Dependence on sex.* Pyle's averages show that the girls make more correct substitutions than the boys at every age from 8 to 18, with a single exception (age 18, digit-symbol test). In the three test sheets the Cincinnati girls are slightly superior to boys in index, while the sex differences in accuracy are too small and inconsistent to be significant, so that speed is the important factor in the better index of the girls. With the 4th (memory) sheet, there is no difference in index at 14, but the girls are superior at 15. Girls at 15 also slightly surpass boys in accuracy on the 4th sheet.

(4) *Dependence on race.* B. T. Baldwin tested 37 white and 30 negro girls at a Pennsylvania Reform School for 16 practise days, 5 min. per day, after eliminating 3 whites and 14 negroes who failed to attain 50 per cent. accuracy. Table 86

[3]Consult Woolley and Fischer for table showing norms of accuracy and for numerous graphs of distribution for the substitution index in relation to school grade.

TABLE 85

Substitution Index, in Sec., Cincinnati Working Children (Woolley and Fischer).

AGE	RANK	SHEET 1		SHEET 2		SHEET 3		SHEET 4	
		Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
14	Best	71.0	83.3	68.5	68.0	59.4	59.8	53.0	52.3
	75th Perc.	147.0	142.0	115.7	108.7	97.4	94.0	89.6	88.5
	50th Perc.	172.7	162.6	133.2	130.4	115.9	112.7	111.2	112.6
	25th Perc.	200.9	185.6	157.6	154.0	138.4	134.8	148.3	150.5
	Worst	400.0	419.5	378.0	298.4	276.4	242.9	1,257.6	525.4*
15	Best	82.2	98.4	55.0	67.4	54.0	59.2	52.6	50.6
	75th Perc.	137.3	130.4	104.6	103.6	92.0	91.9	84.4	85.3
	50th Perc.	157.3	148.6	123.7	119.0	110.7	108.7	104.9	103.9
	25th Perc.	179.0	171.6	145.7	138.1	133.4	128.7	145.7	139.3
	Worst	286.6	307.8	241.3	295.1	355.5	248.6	906.5	19,875.0

*To which should be added one case of complete failure—accuracy only 6 per cent. and index 60,000. In comparing this table with the original text it should be noted that I have reversed the designations of the percentiles, so that 100 per cent. here would represent the quickest performance (smallest index).

shows clearly the superiority of the whites. In general, the negroes make only 62.4 per cent. as many substitutions and 245.3 per cent. as many errors as the whites. The fact that the

TABLE 86

Average Number of Substitutions Made by 37 White and 30 Negro Girls in a Pennsylvania Reformatory (Baldwin)

TRIAL	1	2	3	4	5	6	7	8	9
Whites	23.8	42.6	46.7	54.2	61.7	64.9	67.8	78.3	79.6
Negroes	22.6	27.6	31.2	35.8	46.9	48.0	53.9	57.7	61.5

TRIAL	10	11	12	13	14	15	16	AVER.
Whites	86.9	85.9	89.5	94.1	93.7	100.1	116.5	72.3
Negroes	64.7	76.6	71.6	76.0	78.1	72.3	89.0	55.8

average age of the whites is somewhat greater (16.7 vs. 15.1 years) by no means accounts for these differences. Baldwin notes that there are also distinct qualitative differences in the work of the two races: negro girls are slower to warm up to the task, and first to drop back and lose interest: they cannot be forced or stimulated easily, except temporarily through flattery: their work is more irregular, more subject to moods, less accurate and less neat. "They are partially occupied with the task in hand and partially with a random activity, which consists in mumbling, grumbling, humming or saying original and funny things. This second attitude seems a common trait with the race unless consciously inhibited." [4]

(5) *Dependence on practise.* (*a*) Practise-effects within the single trial of the substitution test are revealed, of course, by comparison of the rate and accuracy of the work in the different sections or sheets into which the material is divided. With Form A or Form B the increase in speed in the 8th as over the 1st section amounts to some 10 to 20 per cent. This improvement is not acquired uniformly, however, from section to section. On the contrary, as Tables 81 and 82 show, there is a tendency toward a decrease of efficiency at about the middle of the work. Thus, in the individual tests both men and women, taken collectively, show a reduction of speed in the 5th section: similarly, in the group tests both men and women write fewer symbols in the 5th than in the 4th 30-sec. period.

In the individual tests, the 4th section comes at the bottom of the first column, the 5th at the top of the second column. The brief delay occasioned by the necessary readjustment (of paper, pencil, attention, etc.) may explain a part, but only a part of the reduction in time.

A plausible explanation is that reported by one *S*, who noted that, in Section 4, being so far from the circles, she relied upon her memory, whereas in Section 5, the very proximity of the circles tempted her to glance at them to make sure of her work, and thus to work more slowly.[5]

---

[4]Since the above was written, Pyle (9a) has published the results of an investigation in the public schools of Missouri which discloses a similar inferiority of negroes to whites in the substitution test. Speaking in general terms, the negroes are less than half as efficient as the whites in the test.

[5]Form A has been devised especially to avoid the variation in distance of test-blanks from the pattern at different periods of the work.

To test this hypothesis, trials were made with 10 college students, using material of Form B, but rearranged (as suggested above) to resemble Form A (the test-blank in one long column sliding beneath the cardboard). The average scores, in sec., per 5-line section, were 55.7, 48.6, 44.0, 40.9, *43.0*, 40.3, 41.5, and 40.8, for the Sections 1 to 8, respectively. (Total time, 354.8 sec.; symbol score, 8.2). There is, then, still a loss of more than 2 sec. at Section 5.

It would appear, therefore, either that the test-material of Section 5 happens to be more difficult than that of Sections 4 and 6, or that, as a final possibility, the slower rate in Section 5 is merely an expression of a mental condition—fatigue, weariness, loss of initial enthusiasm. That this explanation may be entertained is shown in Table 87, where it will be seen that, although more *S*'s lose speed in the 5th than in any other section, there are, nevertheless, numerous instances of loss of speed in other portions of the work, especially in Section 7. The *S*'s of Table 87 are the 10 just mentioned, and the 18 of Table 82.

TABLE 87

Substitution Test. Distribution of Gains and Losses in Speed (Whipple)

SECTIONS	1–2	2–3	3–4	4–5	5–6	6–7	7–8
Number gaining speed	24	17	16	9	19	12	16
Number losing speed	4	7	8	16	8	15	8
Number maintaining speed	0	4	4	3	1	1	4

(*b*) Special investigations upon practise in this test have been made by Starch, Lough, Kline and Miss Munn. Starch's work, which is confirmed by Miss Munn's, shows that relatively short, distributed practise periods are the most effective (Fig. 69); from 10 to 20 min. seems to be best, at least for adults. Lough found no evidence of plateaus in the curve of improvement in tests lasting from 20 to 90 days. Miss Munn found the typical curve of improvement to be rapid in rise at first, then slower. Children were slower at the start, but gained more, absolutely, than did adults. Curves from two aged *S*'s were similar to those obtained from the young. Retrials showed that fairly strong practise-effects persisted for at least as long as 5 mos.

Kline tested the effect of practise in one form of substitution upon performance in other forms of substitution and found that "practise in writing digits for letters is transferred with favorable effect to subsequent work in writing symbols for digits, but is transferred with unfavorable effect to subsequent work

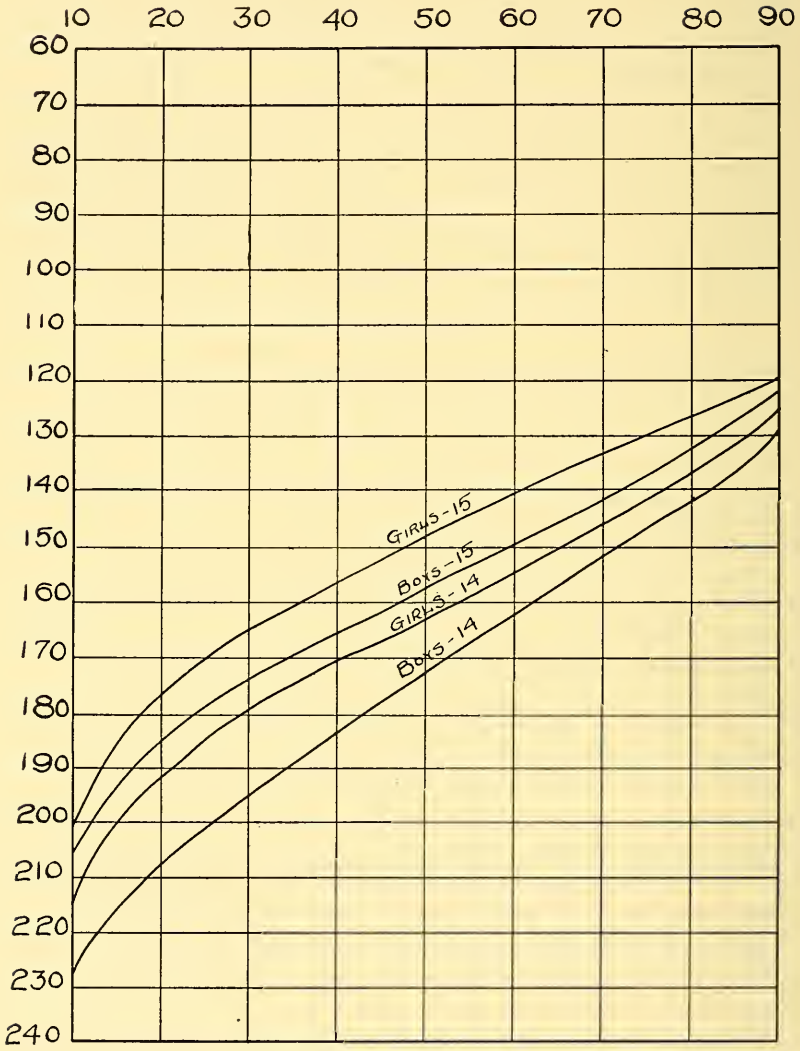

FIG. 67. PERCENTILES FOR THE SUBSTITUTION INDEX FOR CINCINNATI WORKING CHILDREN 14 AND 15 YEARS OLD—SHEET 1
(After Woolley and Fischer)

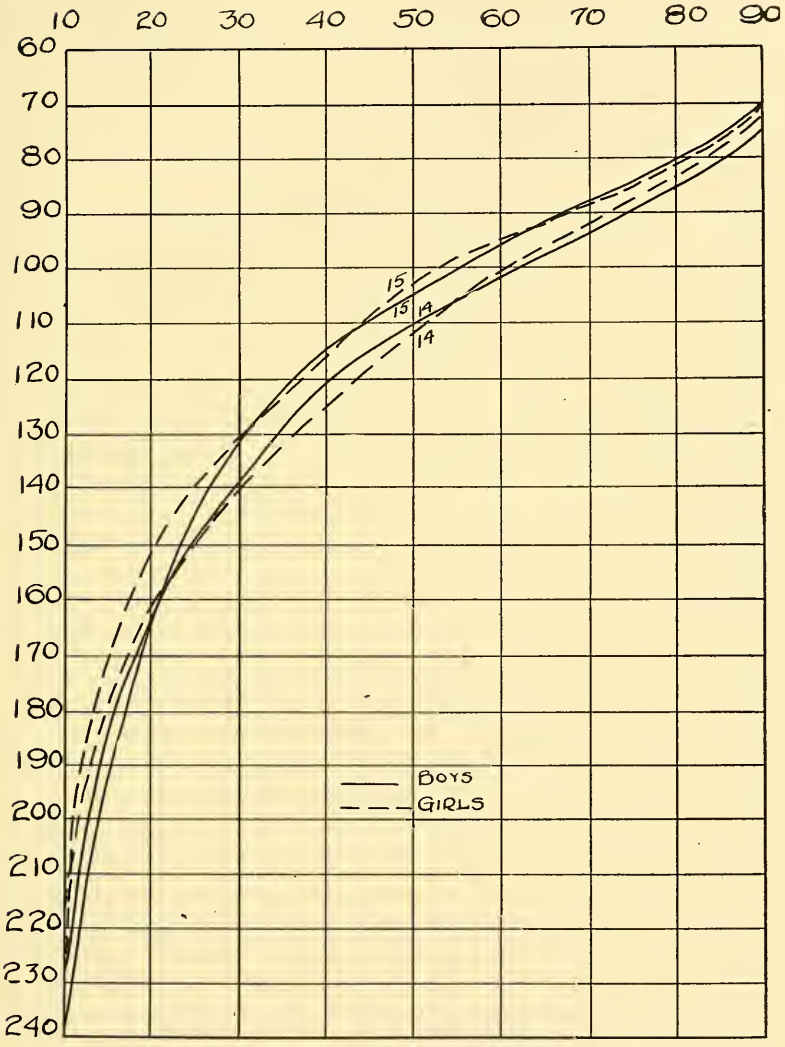

FIG. 68. PERCENTILES FOR THE SUBSTITUTION INDEX FOR CINCINNATI WORKING CHILDREN 14 AND 15 YEARS OLD—SHEET 4 (After Woolley and Fischer)

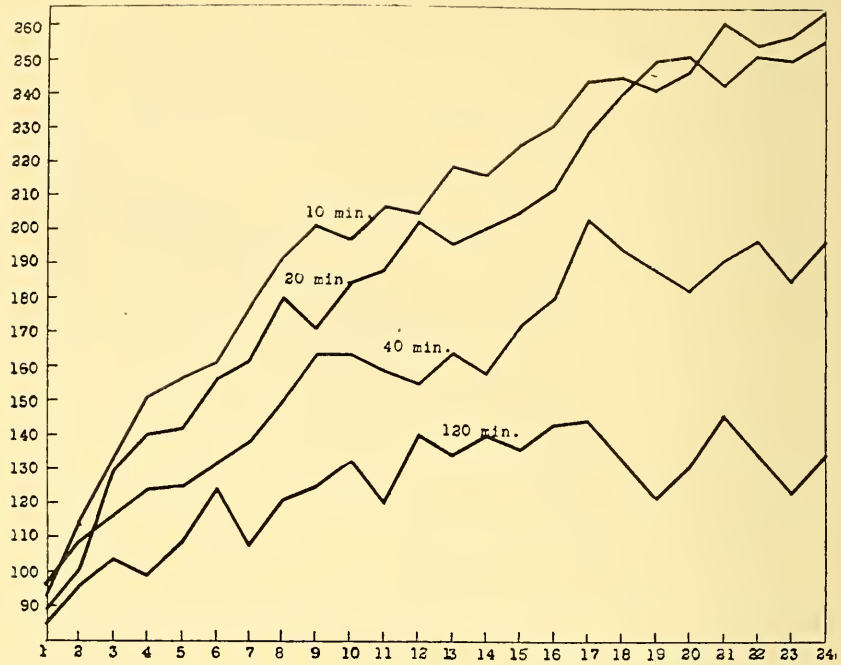

FIG. 69. EFFECT OF PRACTISE PERIODS OF DIFFERENT LENGTHS UPON PERFORMANCE IN THE SUBSTITUTION TEST (From Starch)

Results based on the work of 42 college students. Units on the base-line represent number of successive 5-min. periods: ordinates represent number of substitutions made in 5 min. Designations attached to each curve indicate the length of the work periods of each group.

in writing digits for symbols." Moreover, the more the drill work is spread out in time, the greater, on the whole, is this interference effect. Here the interference is due, of course, to the fact that in the second test-series the same characters must be written as in the drill series, but with different associative connections.

(6) *Dependence on intelligence.* If we admit that the school grade reached by children 14 or 15 years old affords a good indication of their general intelligence, it follows that the substitution test correlates well with general intelligence, since there was found at Cincinnati a "positive correlation with school grade for all four pages of the test, for both sexes, and at both ages" (15, p. 153). This correlation is less evident with

the 4th (memory) sheet, where individual differences are most manifest. It follows that children who have reached higher school grades exhibit a somewhat higher performance in the memory test after having spent a decidedly shorter time in the process of learning. The above results pertain to the substitution index (Form C): as to accuracy, that is also positively correlated with school grade, though not so markedly nor so consistently as the index. Similarly, Woolley and Fischer report that some of the *S*'s sent to them under suspicion of mental deficiency were able to attain fair success with the 4th sheet, but "required a far longer time, often more than twice the time, to reach the result" (p. 244).

The author has compared 6 dull and 5 bright grammar-school boys and obtained the results shown in Table 88. Since, with a single exception, all the dull boys are older than the bright boys—on the average about 3 years older—the actual difference in capacity between the two groups that is ascribable to intelligence is much greater than appears from the averages obtained.

TABLE 88

Substitution Test. Bright and Dull Boys. Form B. Individual Method (Whipple)

	BOY	AGE	SCHOOL GRADE	TIME IN SEC.	SYMBOL SCORE
Dull Group	H.	16:9	7, II	700	6
	K.	13:1	5, II	742	8
	N.	14:9	6, I	422	9
	M.	12:8	6, I	975	1
	B.	12:6	7, II	707	4
	S.	15:2	6, I	660	9
Average		14:2		701	6.17
Bright Group	Br.	11:11	8, I	677	9
	Hu.	12:8	6, II	597	5
	Id.	10:9	6, II	566	9
	Tr.	10:4	6, II	648	4
	Fe.	10:8	6, II	591	9
Average		11:1		615:8	7.2

(7) *Dependence on physical condition.* Miss Kohnky used the Cincinnati substitution test among other tests with pupils of two 5th grade classes in that city in her study of the effects of dental treatment upon physical and mental efficiency. The test was given in October to pupils in Room 18 and Room 21, two comparable groups. The pupils in Room 21 were then given elaborate dental treatment, those in Room 18 were given none. Both groups were retested in the following May. The score for the untreated room was 201.6 for the 1st sheet in October and 110.5 for the 4th sheet in May, a total gain of 91.1: the score for the treated room was 206.6 for the 1st sheet in October and 104.9 for the 4th sheet in May, a gain of 101.6, from which it is argued that the pupils subject to dental treatment developed greater 'learning power.'

(8) *Delinquents.* Miss Weidensall tested 88 Bedford Reformatory women and also a group of Vassar College dormitory maids with the material of Form C. She found that the various reformatory groups differed more from one another in speed than in accuracy, that both the college maids and the reformatory women differed from the Cincinnati working girls more widely on Sheet 1 and Sheet 4 than on Sheets 2 and 3, from which it may be inferred that the working girls make a quicker adjustment to the task and reach a higher accomplishment in distinctly less time. The work with the first and fourth pages divided the reformatory women into two distinct groups which correspond with, and confirm the school's estimate of their intelligence: thus, when the women are divided into two groups, 55 per cent. of the below-grade (schooling less than Grade 5B) are as poor in index of substitution for Sheet 4 as the poorest quarter of the grade group. Again, when the women are divided into smaller groups on the basis of years of schooling, there appears a positive correlation with amount of schooling that is as close as that found at Cincinnati. Reformatory women that had reached the 8th grade in schools were better than 15-year-old Cincinnati working girls in both accuracy and time on Sheet 4, but elsewhere the reformatory women were quite generally inferior to the working girls, and the inferiority becomes increasingly great as the grade at which the reformatory women

left school becomes lower. A general idea of the inferiority is given by Table 89. Finally, the correspondence between the results of the test and general ability is further indicated by a correlation of 0.48, P.E. .06, between rank on Sheet 4 and native ability as estimated by the principal of the Reformatory Industrial School.

TABLE 89 .

Differences, in Sec., Index of Substitution, Form C, between Bedford Hills Reformatory Women and Cincinnati Working Girls 14 and 15 Years Old (After Weidensall)

PERCENTILE	SHEET 1			SHEET 4		
	25th	50th	75th	25th	50th	75th
14 Years	−59.4	−10.5	+6.0	−184.5	−44.9	− 8.1
15 Years	−73.4	−24.5	−6.5	−195.7	−53.6	−11.3

Plus sign indicates that the Bedford group is faster, minus sign slower than the Cincinnati group with which it is compared. The order of percentiles is here reversed from that used in the original tables: here the 75th is better than the 50th percentile.

NOTE.—*S*'s who make the fastest records commonly employ the scheme of holding in mind the entire 5-place number (in Form B), and writing down the symbols while keeping the eyes directed upon the circles. The material in Form A lends itself less easily to this scheme.

REFERENCES

(1) B. T. Baldwin, The learning of delinquent adolescent girls as shown by a substitution test. *JEdPs*, 4: 1913, 317-332.
(2) D. F. Carpenter, Mental age tests. *JEdPs*, 4: 1913, 538-544.
(3) W. F. Dearborn, Experiments in learning. *JEdPs*, 1: 1910, 373-388.
(4) C. T. Gray, A new form of the substitution test. *JEdPs*, 4: 1913, 293-297.
(5) L. W. Kline, Some experimental evidence in regard to formal discipline. *JEdPs*. 5: 1914, 259-266.
(6) Emma Kohnky, Preliminary study of the effect of a dental treatment upon the physical and mental efficiency of school children. *JEdPs*, 4: 1913, 569-578.
(7) J. E. Lough, Plateaus in simple learning. *PsBu*, 9: 1912, 87-88.
(8) Abbie F. Munn, The curve of learning. *ArPs(e)*, 2: 1909, 36-52.
(9) W. H. Pyle, The examination of school children. New York, 1913. Pp. 70, especially 18-22.

(9a) W. H. Pyle, The mind of the negro child. *School and Society*, 1: 1915, 357-360.
(10) Carrie R. Squire, Graded mental tests. *JEdPs*, 3: 1912, 363-380, 430-443, 493-506, especially 432-4.
(11) D. Starch, Periods of work in learning. *JEdPs*, 3: 1912, 209-213.
(12) D. Starch, Experiments in educational psychology. New York, 1911. Pp. 183, especially ch. 7.
(13) Jean Weidensall, The mentality of the criminal woman. To appear in *EdPsMon*.
(14) R. S. Woodworth and F. L. Wells, Association tests. *PsMon*, 13: 1911 (No. 57). Pp. 75, especially 53-55.
(15) Helen T. Woolley and Charlotte R. Fischer, Mental and physical measurements of working children (Studies from the laboratory of the Vocation Bureau, Cincinnati, Ohio). *PsMon*, 18: 1914 (No. 77). Pp. 247, especially 148-184.

TEST 38

Memory for serial impressions: 'Rote' memory.[1]—The essential idea in the several forms of memory test treated under this title is to present a series of discrete impressions (*e. g.*, letters, digits, words), which is, if possible, to be reproduced in correct order and exactly as presented. These tests are to be contrasted with the so-called tests of 'logical' memory, in which the material presented is a logically connected whole, and in which the requirement is to reproduce the substance, or the meaning, of what has been presented. In either test, the reproduction may be immediate or delayed, and the mode of presentation and method of measuring efficiency may be varied in many ways.

Memory for a series of discrete impressions has been used to study individual differences, as conditioned by sex, age, mental ability; to detect fatigue; to investigate the nature of practise, the possibility of training retention and recall, the most economical methods of learning, etc.

To understand the results and conclusions of the small army of investigators of memory, it is convenient to classify the methods and the materials that have been most commonly used.[2]

---

[1]The author desires to acknowledge the assistance of Dr. L. R. Geissler in the collation and sifting of the literature bearing upon this test.

[2]For more extended discussion of the historical development of the several experimental methods, together with accounts of the results that have been attained, the reader should consult Bentley, Binet (9), Burnham, Ebbinghaus, Gamble, Henri, Offner and Pohlmann. The last-named gives a particularly valuable summary of the methods.

CLASSIFICATION OF METHODS FOR MEMORY TESTS

(1) The *method of complete memorization*, or method of complete mastery (*Erlernungsmethode*), developed in the classic work of Ebbinghaus (*Ueber das Gedächtnis*) in 1885, and refined by Müller and Schumann, demands that *S* repeat the series of impressions again and again until he can reproduce it without error, without hesitation, and with certainty of correctness. Efficiency is measured by the number of presentations required for this complete learning.

In practise, this method is frequently supplemented by testing the number of presentations of the same series that is needed to relearn it at any assigned time after the first learning (*Ersparnisverfahren* or *Ersparnismethode*), in which case the saving in number of repetitions in the relearning, as compared with the learning, measures the amount of retention, or the degree to which the first impression has persisted.

(2) The *memory-span method* (*Methode der Gedächtnis-Spanne*), first devised by Jacobs, elaborated by Ebert and Meumann, and extensively used in England and America, consists in the determination of the maximal length of a series of impressions that can be reproduced with a given degree of accuracy (usually complete accuracy) after a given number of presentations (usually, though not necessarily, one presentation). Ordinarily, *E* begins with a series that is easily within *S*'s limit, and increases the length of the series, keeping other factors constant, until errors appear.

(3) The *method of retained members* (*Methode der behaltenen Glieder*), first so designated by Ebbinghaus, but more carefully studied by Pohlmann, consists in the determination of the degree of mastery (proportion of elements correctly reproduced) of a series of a given length, after a given number of repetitions. The method is somewhat like the span method, but the length of the series is so chosen that *S* cannot attain complete mastery. In practise, many span tests actually become tests of degree of mastery.

(4) The *method of right associates* (*Treffermethode*), proposed by Jost and developed by Müller and Pilzecker, consists in presenting a series of impressions (typically, nonsense syllables in trochaic rhythm), and of subsequently testing *S*'s ability to name the member that follows any given member. Usually the accented member is given, and *S* tries to designate the 'right associate' for it. (When his time of response is measured, the method is known, in full, as the *Treffer- und Zeitmethode*). Its special value is to afford opportunity for analyzing the nature of the associative connections; it has not been proposed as a test of efficiency.

(5) The *method of prompting* (*Methode der Hilfen*), somewhat similarly, tests the nature and strength of the individual associative connections in the series, and is of questionable usefulness for practical testing. As illustrated in the work of Ephrussi, the method consists in an attempt by *S* to reproduce the series before it has been fully learned, and in promptings by *E* at each point of hesitation or error. Efficiency is inversely related to the number of promptings required.

(6) The *method of interference of associations* is exemplified in Bergström's study of card-sorting (5). Here 80 cards are sorted by *E* into 10 piles, and subsequently, at a given interval, into another 10 piles differently arranged. The second sorting is slower because of the persistence of associative connections developed in the first trial. Analogous tests can be fashioned with other forms of material, as has been suggested in the Substitution Test.

(7) The *method of reconstruction*, used by Münsterberg and Bigham with colors, and by Miss Gamble with odors, consists in presenting a series of stimuli in a definite order, and then, after a predetermined interval, in presenting the same stimuli in chance order. *S* attempts to rearrange them in the original order.

(8) The *method of recognition* consists in the presentation of a limited number of impressions, which are subsequently presented again, in conjunction with other stimuli, to see how many of the first series *S* can recognize in the second series. Examples will be found in the work of Smith and of Henri.

(9) The *method of identical series*, as employed by Reuther is a modification of the method of recognition, in which the original series is always actually presented intact, though, of course, this fact is concealed from *S*.

(10) The *method of continuous lists* (*Methode des fortlaufenden Niederschreibens oder Aufzählens*), employed by Kræpelin, is identical with the procedure described in Test 33, though sometimes *S* is required to write words that belong to specified categories.

(11) The *method of chance verbal reactions* (*Methode der zufälligen Wortreactionen*), well illustrated by the investigations of Aschaffenburg and G. E. Müller, is the stock association experiment, with emphasis upon the qualitative as well as the quantitative study of the associative sequences. (See Test 33A.)

(12) The *method of description or report* (*Aussage*) is a form of memory investigation with peculiar problems of its own, as has been shown in Test 32. In it, the terms in which the reproduction takes place are not restricted to a direct equivalence with the material presented, but are merely indicative or descriptive of this material.

The tests which follow are primarily intended to test capacity for immediate reproduction after a single presentation, either by the memory-span method or by the method of retained members (degree of mastery). The capacity which is tested corresponds to what the Germans call *Merkfähigkeit*—a term which is perhaps best rendered in English as *immediate memory*. Tests of capacity to recall or to recognize after an interval of greater or less duration would doubtless more nearly measure memory in the more exact sense of that term, but, unfortunately, little attention has been paid to this phase of mental testing, owing presumably to the desire to complete observations in a single sitting.

CLASSIFICATION OF MATERIAL FOR MEMORY TESTS

The material used in tests of serial memory may be classed according to the sense-department to which it is presented (visual, auditory, visual-auditory, etc.), and according to its

nature or form. Again, visual material of different forms may be presented either simultaneously or successively.

(1) *Actual objects* were used by Netschajeff, Lobsien and Kirkpatrick. Thus, Lobsien showed 9 objects at the rate of 1 per sec., *e. g.*, newspaper, key, handkerchief, glass, slate, box, book, glove, chalk. Netschajeff used 12, Kirkpatrick 10 objects.

(2) *Pictures of objects*, 10 in number, were used by Miss Calkins; groups of 20 pictures by Mrs. Squire and by Carpenter, following the suggestion of the earlier Binet-Simon tests.

(3) *Sentences* also form a portion of the Binet-Simon tests, and have been tried by Ritter, Miss Sharp, Mrs. Squire, Carpenter and Abelson. Directions for their use will be found below.

(4) *Words* may be used in the most varied kinds of series. Thus, series of Latin-German, or English-German, or other pairs of nouns, have been used to produce a 'vocabulary' form of test, as by Wessely. A distinction may be made between 'related' or 'associable' terms and 'unrelated' or 'dissociable' terms (Norsworthy: Bergström, 6). For example, *paper, writing, compose*, etc., *vs. horse, bricks, soldier, acorns*, etc. Meumann (51), Burt (16) and Pyle have compared the span (3 to 8-term series) for concrete nouns, *e. g., stove, ink, lamp, street*, etc., with the span for abstract nouns, *e. g., influence*, etc. Netschajeff and Lobsien tested the relative reproducibility of words (12 and 9-term series) that connoted visual, auditory, tactual and emotional ideas, respectively. (Examples: *lightning, dial, sunbeam; thunder, crash, whistle; cold, soft, smooth; hope, doubt, regret.*) Kirkpatrick and Calkins also used 10-term series of words that related to objects, as did Pohlmann. Hawkins compared simultaneous and successive exposure of 15 nouns. Binet, Ritter, Simpson, Abelson, Lapie and Sharp also employed lists of words of varied length and complexity.

(5) *Nonsense syllables* were tried but discarded by Jacobs, likewise by Cohn and Dieffenbacher. They formed, however, the stock material in Ebbinghaus' pioneer work, and were subsequently made more serviceable by the precise rules that Müller and Schumann formulated for their construction. Bergström, Burt, Smith, Müller and Pilzecker, Pohlmann, van Biervliet, and others have found them of value: indeed, Pohlmann contends that, on account of their equivalence one to another and their relative freedom from varying associations in different *S*'s, nonsense syllables form the best and most reliable material for memory tests. Series specially adapted for English readers will be found in Test 25.

(6) *Letters* (usually consonants only, to avoid the formation of syllables or words) have been used by Jacobs, Binet (8), Cohn, Pohlmann, Sharp, Finzi, Smith (71, 73), and Winch (80). An idea of the great variety of procedure that may be developed with a single form of material may be gained by noting that Binet used 15 consonants exposed visually and simultaneously, for 20 sec.; Cohn exposed 12 consonants arranged in the form of a square for 25 sec.; Pohlmann read 10 consonants to his *S*'s 3 times over; Sharp exposed 12 letters successively with the Jastrow drop-apparatus, at the rate of 1 per sec., and repeated until the series was learned; Smith exposed 12 consonants simultaneously for 10 sec., and read other series of 4, 5, 6, 7, and 9-term consonants; Winch repeated 12 consonants auditorily in 25 sec., and also used the letter-square method (described below), as did Wyatt and Anderson.

(7) *Two-place numbers*, administered orally, were used by Schuyten (8 numbers repeated by *S*'s in concert), Lobsien (9 numbers), Pohlmann (10 numbers given three times), and Netschajeff (12 numbers).

(8) *Digits*,[3] *i. e.*, one-place numbers, have been employed by Jacobs, Johnson, Bolton, Binet, Ebbinghaus, Hawkins, Ritter, Chambers, Kohnky, Lapie, Sharp, Smedley, Krueger and Spearman, Wissler, and many others, in the most varied manner (4 to 10-place series, given auditorily, visually —either simultaneously or successively—or in combined appeal to vision and audition, to vision, audition and 'hand' memory, or to vision, audition, and 'articulatory' memory). Abelson appears to be the only investigator to have discarded digits as inappropriate for mental tests.

(9) *Geometrical drawings* have been used by Münsterberg and Bigham, and by Bernstein and Bogdanoff, who selected forms that would be unfamiliar to their *S*'s.

(10) *Lines* of varied lengths have been employed by Toulouse and by Binet (9).

(11) *Miscellaneous visual characters*, symbols, combinations of dots, lines, etc., formed a portion of the material in the investigation of Ebert and Meumann.

(12) *Sounds*, such as those produced by tearing paper, whistling, stamping, ringing a bell, etc., were arranged in 9-element series by Lobsien, and in 12-element series by Netschajeff.

(13) *Memory for commissions* forms a well-known part of the Binet-Simon tests. An extension of this idea into a sort of memory-span test of memory for commissions has been used by Abelson in the study of backward children.

Aside from these wide differences in general method and in form of material, attention should be called to differences in rate or tempo at which the series is first presented, to differences in the number of times the series is presented, and to differences in the time-interval elapsing between presentation and reproduction.

As a rule, the rate of presentation has been not slower than 1 impression in 2 sec., and not faster than 2 impressions in 1 sec. A rate of 1 impression in 0.75 sec. has been found well adapted for adults.

The typical span test is one in which the series is presented but once: from the point of view of functional testing, therefore, the repetition of the stimulus series may be regarded as a variant method, not to be introduced save for the special purpose of studying its effect.

Similarly, as has already been said, the greater portion of the tests here reported have been made with no interval between presentation and reproduction. It is to be noted, however, that Smedley, in his tests of Chicago school children, separated presentation and reproduction by an interval of 5 sec. Wyatt caused his *S*'s to count backward from 20 before writing

---

[3]Reuther has formulated rules for the construction of test-series of digits, analogous to the rules of Müller and Schumann for test-series of nonsense syllables. The following are the most important of Reuther's principles: (1) Do not repeat a digit in the same series (impossible to avoid, of course, in 10-place series). (2) Do not begin a series with the number 1. (3) Avoid the use of zero. (4) Do not place any two digits in their natural relations with one another. (5) Do not use sequences that suggest historical dates. (6) Do not use in immediate succession two series that have the same digit in the same place at any point in the series.

nonsense syllables and introduced an interval of 5 sec. in his tests with letter-squares. Kirkpatrick, and Calkins in her repetition of his tests, secured a reproduction both immediately after, and 3 days after the presentation, in order to contrast 'immediate' with 'delayed' memory or recall. Somewhat similarly, Binet, and Sharp in her repetition of his tests, secured a reproduction of each of seven 7-place word-lists directly after its presentation, and a 'recapitulation,' in so far as it was possible, of the 49 terms at the close of the whole test, *i. e.*, about 3 min. after the first presentation. Binet contrasts, in this way, immediate memory with what he terms 'memory of conservation.'

Since, as the results that follow show, even minor variations in the conduct of a memory test affect its outcome, it follows that the results of different investigators may not be expected to exhibit complete accordance with respect to the relative influence of sex, age, mental ability, etc.

Five chief forms of test have been selected and are recommended as standard for this field of investigation; variant methods are suggested in each case. By reference to the classification of methods and materials just given, *E* can devise further modifications to suit special requirements. These five forms are (1) tests with digits, resembling in scope Smedley's Chicago tests, but with several differences in procedure, (2) tests with letters, after Cohn's method, (3) tests with lists of words, after the methods of Meumann and of Burt, (4) tests with sentences of graded difficulty, and (5) tests with pictures of objects.

A. MEMORY SPAN FOR DIGITS

MATERIALS.—Printed test-cards, 42 in number, arranged in three sets of 14 cards each, for presentation by 3 different methods. (Each set contains 2 cards each of 4, 5, 6, 7, 8, 9, and 10 digits.) Metronome. [For serial visual exposure, in addition, Jastrow's memory apparatus (Fig. 70). Cardboard. Willson's gummed figures, black, Size 5. For letter tests, full sets of gummed letters, Sizes 5 and 10.]

PRELIMINARIES.—On the back of each card write the digits that are printed on its face: this enables *E*, when the test demands it, to pronounce the test numbers while displaying the card to *S*. The purely auditory and the auditory-visual-hand-motor series are not included in the printed cards, but should be prepared by *E*, preferably, for convenience, on a single piece of cardboard, the size of the printed cards. For the auditory series, use the following numbers, in the order given: 6135,

2947, 36814, 57296, 241637, 935816, 8537142, 9412837, 47293815, 71836245, 924738615, 475296318, 8697132504, 2146073859. For the visual-auditory-hand-motor series, use these numbers, reversed, *e. g.*, 5316, etc.

METHOD.—If only a single test can be made, employ the visual-auditory-articulatory form of presentation, since this is most likely to produce uniform conditions of ideational imagery for all *S*'s. But if the tests can be taken in full, follow the order of presentation outlined herewith.[4] In any event, preface each form of presentation with a special, short sample-series, without demanding reproduction, in order that *S* may be perfectly clear as to the nature of the test. Within each form of test, also, preface each presentation with a statement of the number of members in the coming series, *e. g.*: "This will be a series of 5 digits." The metronome should be set at 60, *i. e.*, one stroke per sec., for all tests.[5]

(1) *Auditory presentation.* Explain the test by a simple illustrative series. Require *S* to close his lips firmly, and to press his tongue against the roof of his mouth—this to reduce the tendency to articulation, and in group tests (all of the memory tests lend themselves well to group presentation) to avert communication between *S*'s. Start the metronome.[6] Pronounce the digits, one at a time, with the utmost care to ensure

[4]It goes without explanation that the longer series may be omitted with very young, the shorter with mature *S*'s. Use, for the shortest series, one that is easily within the span of the poorest *S* to be tested, for the longest series, one that is too difficult for the best *S* to reproduce without error.

[5]It may be well at this place to point out the differences between this procedure and that followed by Smedley at Chicago. Smedley used no series longer than 8. He gave no warning of the length of the coming series. He set the metronome at 90. He did not present the several series in regular order, but irregularly, though beginning with an easy series. He inserted an interval of 5 sec. between presentation and reproduction. He distributed his tests, seven in all, at hourly intervals. Finally, he gives no clear statement of his method of computing results, save that the "percentage correctly recorded constituted the grade."

[6]If he finds it necessary, *E* may substitute a silent metronome, made by swinging a small weight on a string, but the fact that the regular metronome is somewhat noisy should not be taken as evidence that it disturbs *S;* on the contrary, a noise of moderate intensity is not infrequently found to be a stimulus to better attention. Moreover, the ticking metronome is much more serviceable when *S* is asked to pronounce the digits in conjunction with *E*, and it probably operates to some extent to break up tendencies to learn the digits by grouping.

even tempo, clear articulation, and entire absence of rhythm.[7] Directly at the conclusion of the series, let *S* repeat as much as possible of it. Although, under some circumstances (with very young or backward *S*'s), an oral reproduction may be imperative, a written reproduction should be considered standard, both because the proper placing of the digits furnishes *E* with data for scoring *S*'s performance (and the placing must indicate possible omissions), and because experiment shows that, at least for maturer *S*'s, written reproduction is preferred, and is more successful than oral reproduction. *S*'s recall should, therefore, be entered upon a prepared blank, with the caution to indicate every omission by a dash or a blank space.[8]

(2) *Visual presentation.* Use Cards V-4a, V-4b, etc., to V-10b. Follow the directions for auditory presentation, but in place of pronunciation, exhibit the entire card for a length of time identical with that for auditory presentation, *i. e.*, with an allowance of 1 sec. per digit. The metronome should be used here, as in all phases of this test, in order to keep the conditions of presentation comparable. It probably also tends to induce *S*'s to apprehend the digits successively and in the same tempo as that used for auditory presentation. Note to what extent *S* articulates the digits: even with lips and tongue placed as directed, they will often be seen to move, and contractions of throat muscles may also indicate partial articulation.

---

[7]The difficulty of speaking without accent, or without grouping the digits, has led Binet to reject oral, in favor of visual presentation. Even if *E* pronounces without accent or rhythm, there is no guarantee that *S* may not mentally cast the digits into a strongly accented and grouped series, and, in fact, mature *S*'s, working with the longer series, are almost certain to catch this 'trick' in time. Ritter advocates that *E* should give a decided objective rhythm to every series on just this account; this factor will then form a constant, rather than a variable 'error.' One difficulty with this plan lies in the fact that, in using series of varying lengths, it is impossible to use any constant metrical phrasing.

[8]For group work, the class should be provided with blank forms, so numbered and arranged that no misunderstandings may occur on the part of *S* in entering the data, or on the part of *E* in interpreting it. Allow ample time for writing. Netschajeff, Pohlmann and Schuyten all found 2 min. desirable in classroom tests. In group tests, care must be taken to prevent audible repetition of the digits during the reproduction.

(3) *Auditory-visual presentation.* *E* presents the cards, as in the purely visual procedure, but also pronounces the digits, as in the auditory procedure, by reading them from the back of the card. *S* sees and hears the digits. Cards AV-4a to AV-10b are used.

(4) *Auditory-visual-articulatory presentation.* *E* presents the cards as in (2). *E* and *S* pronounce them in concert, in time with the metronome. *S* sees, hears, and pronounces the digits. Cards AVA-4a to AVA-10b are used.

(5) *Auditory-visual-hand-motor presentation.* *E* pronounces the digits as in (1) : *S* writes them, as fast as pronounced by *E*, upon scrap paper: when the series is finished, *S* at once discards the scrap paper, and reproduces the series. *S* hears, sees, and writes the digits. Use the same numbers as in (1), but reverse the order of the digits. In this test, it will ordinarily be necessary to devote one or two preliminary trials to fore-exercise.

VARIATIONS OF METHOD.—(1) Meet the bothersome tendency toward grouping and rhythmizing—bothersome because exhibited by some *S*'s and not by others—by presenting the digits in trochaic rhythm: this device is perhaps favored by selecting series of 4, 6, 8, and 10 digits only.

(2) Introduce a time-interval between presentation and reproduction. If this interval is short, it may with advantage be occupied with some form of distraction, like saying the alphabet in concert, since the effect will be more like that of a much longer 'empty' interval. The disadvantage of an unoccupied interval is that some *S*'s will mentally rehearse the series just presented.

(3) Substitute successive for simultaneous visual presentation in Forms 2, 3, and 4. For this purpose, *E* must prepare cards for insertion in the Jastrow memory apparatus,[9] so that

[9]Jastrow's instrument is adequate if *E* is careful to make the exposures regularly, in time with the metronome; it is especially useful for group tests. If *E* desires a more accurate exposure apparatus, for individual tests, he may employ the Ranschburg memory-apparatus (now improved by Wirth), Kuhlmann's memory-apparatus, Bergström's rather elaborate exposure apparatus, or G. E. Müller's modification of the kymograph for 'step-fashion' exposure, as described, in improved form, by McDougall. Burt, however, contends that the distraction produced in immature and inexperienced *S*'s by the sight of unfamiliar apparatus more than counterbalances the advantage of greater precision, mechanical regulation of rate

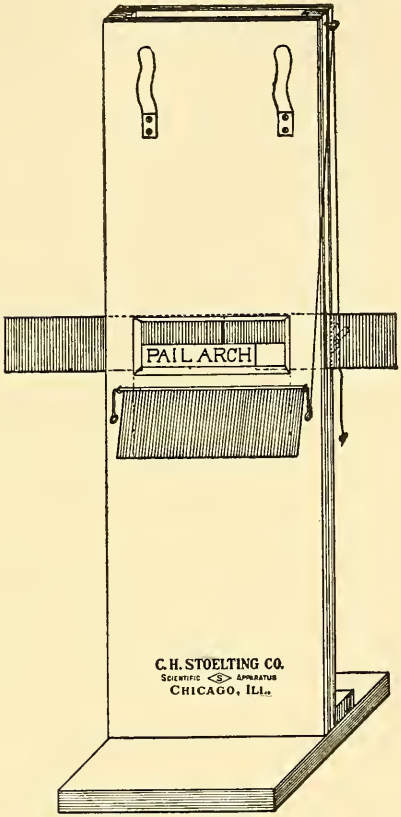

FIG. 70.—JASTROW'S MEMORY APPARATUS.

the numbers used in Forms 2, 3, and 4 (above) may now be exposed in vertical columns. In order to secure sufficiently long series, the exposure-lever of the instrument is so inserted

---

and duration of exposure, etc.; he used, for successive exposure, a slotted piece of cardboard, which was shoved along the column of impressions by *E* (apparently at no uniform or constant rate, but as fast as proved convenient to *S*).

Kuhlman arranges to have each exposure followed by a blank section of perhaps a different duration from that of the exposure. The idea is to control the amount of time that *S* can spend in re-imaging or recalling the impression just received. According to Kuhlmann, *S*'s use from one-half to two-thirds of the total time at their command in ordinary presentations of material for memory tests (whether simultaneous or successive) in this process of re-imaging. The importance of the process, he thinks, varies much with individuals.

as to articulate with the pegs that provide a drop of 1 in. at each exposure. Black letters 7/8 in. high (Willson's, Size 5) may then be used. These are visible to the normal eye at 50 ft., but *E* should take the precaution, in classroom tests, to seat myopic *S*'s near the front of the room.

(4) Test the effect, upon a series too long for *S* to reproduce in one presentation, of two, three, or more presentations in immediate succession.

(5) Give repeated tests by the same method, with a series of a given length (in excess of *S*'s span), to test the effect of practise.

(6) Change the rate of exposure from one impression per sec. to one impression in 2 sec.

(7) Keeping other conditions (form of presentation, length of series, etc.) constant, compare *S*'s efficiency under normal conditions with that under different forms of distraction. Smith (71) used for this purpose three different concomitant activities: his *S*'s were required during the presentation (*a*) to tap in time with the beat of a metronome, (*b*) to repeat the syllable *la,* or (*c*) to add mentally by 2's or by 3's.

(8) Prepare cards with letters[10] in place of digits, for use by any of the procedures above described. Use only consonants. Avoid alphabetical sequences, or suggestions of words or abbreviations.

Treatment of Data.—(1) If it is desired only to determine *S's* memory span, *sensu stricto,* this is indicated simply by the maximal number of digits that can be reproduced without error of any kind.

(2) If, as is more usual in comparative tests, it is desired to determine the degree of correctness with which series longer than the span are reproduced, the simplest plan is to assign arbitrary scores to the various forms of error. Ebbinghaus, for example, scored every omission as 1 error, every displacement from the correct position in the series by 2 or 3 places as 0.5 error, and every displacement by 4 or more places as 1

[10]Use Willson's black gummed letters, Size 5, for the Jastrow apparatus, or Size 10 to duplicate the regular printed test-cards.

error. *S*'s should then be compared with respect to their error-score in series of each length separately.

(3) A more scientific method of determining efficiency is that of computing the degree of correlation between the order of impressions as reproduced by *S* and their order as presented. This is accomplished, following the example of Krueger and Spearman, by applying Spearman's 'footrule' formula for correlation (see Ch. III), though, in this connection, it is better to modify this formula by counting the sum of *all* the deviations between the two series, rather than the sum of all the positive, or of all of the negative deviations.

For treating the data of these memory tests, therefore, the formula may be written:

$$R = 1 - \frac{\Sigma d}{(n^2 - 1)/3}$$

The computation of Σd needs a little explanation. The following cases may be considered:[11]

(*a*) Suppose that *S* reproduces all the terms of the original series, but not in the correct order. The sum of the deviations is then easily computed. In Case A, Table 90, for instance, the sum of the deviations is 6, and since $n = 10$, by the formula just given, $R = 0.82$.

TABLE 90

Use of the 'Footrule' Method in Scoring the Memory Test (Spearman)

ORIGINAL SERIES	CASE A		CASE B		CASE C	
	Reproduced	Deviations	Reproduced	Deviations	Reproduced	Deviations
3	3	0	3	0	3	0
7	7	0	7	0	7	0
9	2	2	-	?	2	2
4	9	1	9	1	9	1
2	4	1	4	1	4	1
1	1	0	1	0	2	?
0	0	0	-	?	0	0
8	5	1	-	?	5	1
5	8	1	8	1	8	1
6	6	0	6	0	6	0
Sum of deviations		6		12.9		9.3

[11]The author is indebted for these illustrations to a personal communication from Professor Spearman.

(*b*) Suppose, Case B, that certain terms have been omitted. The deviations of the terms given are figured as before. There is then added the amount of deviation to be expected for the omitted terms, on the assumption that they are distributed by mere chance. The chance deviation for each term is $(n^2-1) \div 3n$. In Case B, then, there are three omitted terms, each of which deviates by chance 3.3 places. Hence, the Case C, the total deviation $= 6 + 3.3 = 9.3$.

(*c*) Suppose that *S* reproduces certain terms more than once, *e. g.*, the digit 2 in Case C. In this case, the nearer of the two digits is considered as the correct one. The other, or duplicated, term should be regarded as an omission, and treated by the formula just given. Thus, in Case C, the total deviation $= 6 + 3.3 = 9.3$.

(*d*) Suppose that more than the correct number of terms are reproduced: here the superfluous numbers may be ignored, since, save in exceptional cases, they bring about their own penalty by disturbing the correspondence of order.

B. THE METHOD OF LETTER SQUARES

The idea of displaying simultaneously a series of consonants in a simple spatial pattern appears first to have been suggested by Binet and Henri (11): the method was extended by Cohn, who used it to compare the relative values, for a given *S*, of visual and of auditory-motor learning; and it has since been frequently used with modifications (see, for example, Titchener, 77, 396 ff.) as a method of studying ideational types. Winch used the method to compare immediate with delayed reproduction, Smith to compare various forms of distraction, Anderson and Winch to note the relation to sex and age, Wyatt to compare with school standing.

MATERIALS.—A set of 10 printed test-cards. Prepared forms upon which the reproduction is entered. Stop-watch. [The letter-square cards are printed in large type to make the test available for group procedure. The arrangements avoid, so far as can be foreseen, the use of collocations that might serve as aids to memory. Only consonants are used. The blank forms are ruled in sets of 12 squares.]

METHOD.—Explain to *S* the general nature of the test. Inform him of the duration of exposure, but give him no directions as to how he shall attempt to learn the arrangement of the letters. Expose the stimulus card for 25 sec. Let him fill out the blank form immediately after the exposure. Allow 30 sec. for writing. Repeat with other cards, until 4 to 10 trials have been made.

VARIATIONS OF METHOD.—(1) Defer the reproduction for 20 sec. (or 10 sec., to follow Cohn) after the exposure. Direct *S* to count aloud during this interval, from 1 to 20, 1 number per sec., in time with *E* (who may follow a silent metronome swinging once per sec.). The object is to subdue or eliminate the 'memory after-image,' and to secure true recall—in the strict sense of recalling an experience which had not been just previously in consciousness.

(2) Direct *S* to read the letters aloud, twice over, in concert with *E*, at the rate of 1 letter per sec. Read by horizontal lines. Reproduce with or without the 20 sec. interval.

(3) Direct *S* to repeat aloud, continuously and rapidly, during the exposure, the syllable 'Ah.' Reproduce preferably after the 20 sec. interval filled with the counting. This form of procedure obviously favors the visual memory. If more than one trial is made, use other syllables, such as 'La,' 'Oh,' etc., to avoid the lapse of articulation to automatism.

(4) Direct *S* to count aloud by 2's during the exposure (*e. g.*, 2, 4, 6, or 3, 5, 7, etc.) or to count backwards from 20.

(5) After exposure by any of the methods just outlined, point to one square after another of the blank forms, in irregular order, asking *S* to name or to write the appropriate letters as rapidly as possible. Or, without previous warning, ask *S* to fill in the blank squares in vertical rows, or in horizontal rows from right to left. In theory, visual-minded *S*'s can accomplish this without effort, whereas purely auditory-minded *S*'s must retrace their verbal associations to find the necessary letters.

TREATMENT OF DATA.—(1) Following Winch, assign 3 for each letter in its right position, 2 for each letter one remove to the right, or left, or above, or below its right position, 1 for each letter two removes to the right, or left, or above, or below.[12]

---

[12]This method of scoring possibly puts somewhat too much stress upon right position; at least, in cases like the letter L in the specimen it may be felt that it should not go without credit because it is both in the wrong row and in the wrong column. However, the method above described is the one that has been followed by all who have worked with letter-squares as a mental test.

Specimen of test given				*Specimen of a marked paper*			
M	T	D	X	M(3)	T(3)	L(0)	R(1)
V	L	Y	N	L(2)	V(2)	Y(3)	N(3)
S	Z	B	R	Z(2)	B(2)	S(1)	X(1)

Score: 23 out of a possible 36.

(2) If *S* be competent to render introspective accounts of the manner in which each letter was recalled and placed, *E* may, for qualitative purposes, compute separately the score for letters recalled visually, auditorily, or in other ways.

C. MEMORY FOR CONCRETE AND FOR ABSTRACT WORDS

The essential idea of this test of memory, as devised by Meumann, and followed, with some modifications, by Burt and by Pyle, is to compare *S*'s reproduction of a list of concrete, with his reproduction of a list of abstract terms, given under identical conditions.* The comparison is based not only upon the simple quantitative efficiency in the two forms of test, but also, and more particularly, upon the qualitative analysis of the errors in the reproduced lists. Moreover, the test aims to determine not only *S*'s capacity for immediate memory, but also his degree of intelligence, or grade of mental development. The test rests in principle upon two propositions; first, that words whose meaning is understood are more easily retained and reproduced than words whose meaning is not understood; secondly, that progressive mental development implies progressive comprehension of abstract words.

MATERIAL.—For auditory presentation, use the following lists. For visual-auditory presentation, use the same lists printed upon sheets of cardboard with Willson's gummed letters. For visual presentation, serial exposure with the aid of the Jastrow or other exposure apparatus is recommended.

Three-term lists		*Four-term lists*		*Five-term lists*	
Concrete	*Abstract*	*Concrete*	*Abstract*	*Concrete*	*Abstract*
Street	Time	Spoon	Phase	Ground	Tact
Ink	Art	Horse	Work	Pen	Scope
Lamp	Route	Chair	Truth	Clock	Proof
		Stone	Thing	Boy	Scheme
				Chalk	Form

Six-term lists		*Seven-term lists*		*Eight-term lists*	
Concrete	*Abstract*	*Concrete*	*Abstract*	*Concrete*	*Abstract*
Desk	Space	Ball	Craft	Coat	Law
Milk	Creed	Sponge	Myth	Girl	Thought
Hand	Pride	Glass	Rate	House	Plot
Card	Guile	Hat	Cause	Salt	Glee
Floor	Pledge	Fork	Style	Glove	Life
Cat	Cue	Stove	Youth	Watch	Rhythm
		Post	Mood	Box	Faith
				Mat	Mirth

The above lists are prepared with the idea of confining the abstract terms to words of one syllable, as done by Burt and by Pyle (whose lists are quite similar to the above). This restriction materially lessens the difference in difficulty between the concrete and abstract lists. To duplicate Meumann's conditions the following abstract lists may be substituted for those given above:

Four-term list	*Five-term list*	*Six-term list*	*Seven-term list*
Selection	Society	Conscience	Assumption
Analysis	Symbol	Investigation	Recognition
Explanation	Arrangement	Symptom	Origin
Character	Humanity	Formation	Influence
	Theory	Complexity	Development
		Experiment	Organism
			Value

Eight-term list
Behavior
Tendency
Interpretation
Condition
Opinion
Capacity
Profession
Connection

METHOD.—For group tests, follow Meumann's procedure. Explain the nature of the test and give a sample exercise. Provide each *S* with blanks so arranged that his reproductions may be properly recorded, the lists carefully separated, and dashes inserted for all words omitted. Make clear that the lists are to preserve the order of presentation so far as possible. Before each presentation, notify the *S*'s of the number of words to be spoken. Enunciate with great care, and without grouping, at the rate of one word per sec. Instruct the *S*'s to write their

lists immediately after the presentation, and as rapidly as possible, without trying to 'write their very best.' Guard against interruption, intercommunication, or other possible disturbances. Give the series in order, as above, *i. e.*, 3-term concrete, 3-term abstract, 4-term concrete, etc., save for mature *S*'s, for whom the beginning is to be made at the shortest list that all can accomplish, and for whom lists of more than 8 terms may be arranged by combining some of the shorter unused lists.

VARIATIONS OF METHOD.—Consult suggestions for the memory span for digits (Variations of Method, 2 to 7).

TREATMENT OF DATA.—(1) The simplest method is to disregard the question of order and simply to credit *S* one for each word correctly recalled. This scoring was used by Simpson with lists of 16 words, but is not recommended by him on account of its failure to penalize for erroneous insertions.

(2) Another very simple device is that adopted by Pyle of crediting one for each word correctly reproduced, plus one more for each word placed in the right order. It is evident that this scoring is not specific enough to deal adequately with the various possibilities of insertion, substitution, transposition, etc.

(3) Memory for words may be scored by any one of the three methods already proposed for memory span (Treatment of Data).

(4) The second method proposed for the memory span (arbitrary scores for various forms of error) is followed in principle by Burt in his special system of scoring memory for words: each correct word correctly placed counts 4; each correct word misplaced by one move counts 3; each correct word misplaced by more than one move counts 2; omissions or substitutions count 0. Other rules which he followed concern words with slight alterations; these, in the author's judgment, are not important enough to justify their use unless nonsense syllables are used.

(5) A very elaborate analysis of memory for words was employed by Meumann. For a careful scrutiny of the performance in this test *E* may prefer to adopt such a method, following as

a suggestive pattern the schema below, which has been transcribed from Meumann with a few minor modifications.

ILLUSTRATION OF THE TREATMENT OF DATA IN MEMORY FOR WORDS

Subject: Adolph L. Age, 8 years.

	Types of Error	Number
1.	Memory errors (omissions and displacements) concrete lists.	5⅔
2.	Memory errors (omissions and displacements), abstract lists..	7⅔
3.	Insertions	4
4.	Insertions of nonsense words	1
5.	Fusions	0
6.	Perseverations	3
7.	Regressive inhibitions	1
8.	Complete reversals	1
9.	Substitution of synonyms	0
10.	Substitution of concrete for abstract	1
11.	Wrong formations	4
12.	Misunderstood abstract terms	5
13.	Spelling	Very bad
14.	Handwriting	Undeveloped and ugly

(1) and (2) *Omissions* are represented by the integers, *i. e.*, Adolph L. omitted five words from the concrete lists, 7 from the abstract (the test was carried to the 7-term list only). *Displacements* from the correct order count ⅓ error when the displacement is by one remove only, ⅔ error, when more than one remove (save that with younger children, as in the case above, all displacements count ⅓.) Hence Adolph L. made 2 displacements in the concrete, 7 in the abstract series.

(3) *Insertions* are the total number of words added. These are counted as 1 error each, unless the added word has some similarity of sound to a word actually presented, in which case it counts ⅔ error.

(4) This rubric embraces the relatively infrequent addition of a *meaningless word* that has no similarity in sound or spelling to any of those presented.

(5) *Fusions* of two or more totally independent, successive terms into a single meaningless term are a very significant form of error, which appears in abstract lists written by *S*'s of poor intelligence, *e. g.*, *Organ* and *Gattung* are reproduced as *Orgattung*. Mostly found in children 8 and 9 years old.

(6) *Perseverations* are indicated by the recording by *S* in a given series of a word that had already been reproduced in an earlier series. If frequent, this is a sign of a low intelligence, lack of self-control and of critical judgment.

(7) *Regressive inhibitions.* Failure to reproduce at least one-half of the terms given is, as a rule, to be interpreted as regressive inhibition. This condition is commonly attributable to a state of confusion into which a child is thrown, when he is suddenly 'overwhelmed' by the task, when everything 'flies out of his mind,' he 'loses his wits,' and is unable to accomplish even a fraction of his normal performance. The same thing is seen in adults under conditions which are difficult for them. Since, Meumann argues, this is essentially due to inability to force attention, lack of this ability is a token of poor general ability, and hence of low intelligence. Failure due to absolute lack of intent to succeed must, of course, be distinguished from the lack of ability to succeed.

(8) *Complete reversal* of word order, either in a large portion, or in the whole of a list is "a peculiarly puzzling phenomenon." There are occasionally met, for instance, cases in which a series of 8 words are all written in the reverse of the order presented.[13]

(9) The *substitution of synonyms* refers to the easily intelligible cases in which a word of like meaning, but different sound, replaces the word given, *e. g.*, *road* for *street*.

(10) The *substitution of concrete for abstract words* refers to the use of concrete terms of similar sound, whether of similar meaning or not, *e. g.*, *cows* for *cause*, *simple* for *symbol*. *E* must use his judgment here in making allowances for faulty spelling.

(11) Wrong formations, especially the use of wrong endings, constituted a prolific source of error in the German tests, particularly with abstract words, *e. g.*, *Glaubheit* for *Glaube*. Errors of this type may be expected to be less frequent in the less highly inflected and compounded English language, but occasional instances will be found, *e. g.*, *selectness* for *selection*.

(12) *Misunderstood abstract terms* is to be regarded (as the author understands it) as expressing the sum total of misapprehended abstract terms, whether the misunderstanding is indicated by substitutions, faulty endings, fusions, very faulty misspellings, or in other ways.

(13) *Orthography* constitutes a secondary symptom of intelligence. In order to estimate spelling fairly, papers are ranked as 'poor' in spelling only when the sum of misspelled words is 50 per cent. or more greater than the average number of misspellings for *S*'s class.

(14) *Handwriting* constitutes another secondary symptom of intelligence, and is merely rated, as fairly as possible by comparison of numerous papers, as good, average, or poor.

These 14 rubrics are filled out for each *S*. For the estimation of memory capacity, pure and simple, Meumann takes Nos. 1 and 2; for the estimation of intellectual ability, he divides the rubrics into three groups, (1) those that serve as indirect indexes of intelligence (Nos. 1, 2, and 3), (2) those that serve as direct evidence of intelligence (Nos. 4 to 12, including a statement of the relation of Nos. 1 and 2), and (3) those that serve as secondary symptoms of degree of mental development (Nos. 13 and 14). Now, for each of these condensed indexes, the grade of each *S* is indicated as (1) above average, (2) average, or (3) below average, and final comparisons and correlations are based upon these grades.

---

[13]The author is inclined to regard this phenomenon as a simple case of attempt on the part of a few *S*'s to get the series right by beginning with the last word heard and working back to the first section. *S* may have intentionally disregarded instructions to reproduce in the order given, or may have interpreted these instructions to include the reverse order as acceptable. In other words, it scarcely seems probable that the child does not know that he has reversed the order of presentation.

D. MEMORY FOR SENTENCES

In the first (1905) series of tests proposed by Binet and Simon there was included a test of memory for sentences; in the second (1908) series, sentences of 6, 16 and 26 syllables were inserted in the tests for 3, 6 and 12 years, respectively; in the third (1911) series a test of memory for sentences is used at 5 and 15 years. In my previous account of the 1908 series I inserted a provisional set of 21 sentences ranging in length from 2 to 42 syllables. These were subsequently tried out by Mrs. Squire, and another analogous set of 21 sentences was arranged by Carpenter when he repeated Mrs. Squire's test. Abelson also used a set of sentences of progressive length in his test of backward children.

MATERIAL.—Two printed slips each containing 21 test sentences (2 to 42 syllables). [For visual presentation, two pieces of cardboard.]

Set I is the same as that published by the author as an adjunct to the Binet-Simon tests (1908 series), save that the 11th sentence has been made easier and the 12th and 17th sentences have been made harder to remedy the discrepancies found by Mrs. Squire in the original set (74, p. 379).

Set II is the same as that published by Carpenter, save that the 7th and 9th sentences have been simplified to remedy the discrepancies that he pointed out and that the 10th sentence has been replaced by another, because, in my judgment, it differed markedly from the others in content and meaningfulness for children.

No attempt has been made to equate the two sets in respect to difficulty of corresponding sentences. Set II is probably less well-arranged than Set I.

METHOD.—Explain to *S* that he is to repeat, after once hearing, a number of sentences; that these will be given one at a time, beginning with an easy sentence and becoming more and more difficult. Make clear that he must try to repeat the sentence exactly, word for word. Let *S* sit with his back to *E*. Begin with a sentence well within *S*'s grasp—say, with the second sentence for 6-year-old children, or with the fourth or fifth sentence for older ones. Read each sentence but once, slowly and distinctly. Proceed until positive that no more sentences could be correctly repeated, until, say, *S* has failed with three sentences in succession. Failure is recorded for any altera-

tion, even for a single substitution, insertion or omission. Repeat, if desired, with Set II.

VARIATION OF METHOD.—For auditory-visual presentation, arrange the two pieces of cardboard to display the material, one sentence at a time. Let *S* read them aloud, once over, and immediately repeat what he has read.

For the many other possible variations of method, such as increasing the number of presentations, introducing an interval between presentation and reproduction, see under Variations of Method in the preceding tests of rote memory.

TREATMENT OF DATA.—The simplest plan is to treat the test as one of memory span for sentences. *S*'s score would then be the longest sentence that he could repeat without error.

The difficulty which may then arise from missing one sentence and succeeding with the next is perhaps, however, best resolved by using for the score the total number of sentences correctly repeated; thus, if the first 8 are correct, the 9th missed, the 10th accomplished and the remainder missed, the score is 9 sentences, not 8 or 10.

On account of the individual differences in the difficulty presented by the same sentences to different *S*'s, the unreliability of the test should be lessened by using both sets of sentences whenever time permits; in this case the average score secured by the two trials may be taken as *S*'s final record.

E. MEMORY FOR PICTURES OF OBJECTS

The first Binet-Simon series (1905) contained one test (No. 17) in which a card of 13 pictures of objects was shown. Decroly and Degand used 3 sets of 8 pictures each. Mrs. Squire and Carpenter used a card of 30 such pictures, while Lapie presented a series of 8 pictures, successively, at the rate of one per second.

This test, which has an obviously close relation to the test of visual apprehension (No. 25) and to Binet's card of objects (No. 32A), would appear to have some advantages over more formal material, like digits and letters, for use with younger children.

MATERIAL.—Cardboard on which is pasted small colored pictures of 13 familiar objects. Stop-watch.

METHOD.—Explain the nature of the test to *S;* inform him that he is to have half a minute to look at the pictures, and that directly afterward he is to name as many of the objects as he can. For adults or older children it may be found desirable to reduce the time of exposure to 15 sec.

VARIATIONS OF METHOD.—The mere enumeration of the objects may be supplemented by a demand for further description of them, especially of their colors, or of their location on the cardboard. Mature *S*'s may be quizzed concerning their method of memorizing and recalling the pictures. They may also be tested for recognition by presenting the cardboard, after they have named as many items as possible, to see whether the omitted items can be readily recognized, or they may be given a typewritten list of 30 or 40 objects (including those on the card) from which they are to pick those presented. (Cf. Test 25, B and Test 32, A.)

TREATMENT OF DATA.—Score one for each object correctly named. Make record of insertions and substitutions. Note which objects are most often, which least often recalled.

RESULTS.—(1) *Norms of performance* for memory span for digits will be found in Tables 91 to 94, for letter squares in Tables 95 and 96, for words in Tables 97 and 98, for sentences in Table 99 and for pictures of objects in Table 100.

TABLE 91

Norms of Memory Span for Digits, as Conditioned by Age (Smedley)

AGE	AUDITORY SPAN	VISUAL SPAN	AGE	AUDITORY SPAN	VISUAL SPA N
7	5	5	14	6	7
8	5	5	15	6	7
9	5	6	16	6	7
10	6	6	17	7	8
11	6	6	18	6	7
12	6	7	19	7	8
13	6	7			

TABLE 92

Development of Memory for Digits (Smedley)

AVERAGE AGE		NUMBER TESTED	PER CENT. REPRODUCED	
Years	Months		Auditory	Visual
7	8	19	36.4	35.2
8	8	58	44.6	42.8
9	6	100	45.0	47.4
10	5	89	49.4	54.6
11	6	91	55.4	64.7
12	6	93	55.7	72.3
13	7	109	57.9	76.8
14	6	114	66.2	80.5
15	6	94	65.6	78.2
16	6	77	66.9	81.3
17	6	56	65.5	84.1
18	5	25	67.2	77.5
19	5	12	70.0	85.3

TABLE 93

Dependence of Memory Span for Auditory Digits on Age (Jacobs)

AGE	8	9	10	11	12	13	14	15	16	17	18	19
Number tested	8	13	19	36	41	42	42	72	66	50	30	14
Average Span	6.6	6.7	6.8	7.2	7.4	7.3	7.3	7.7	8.0	8.0	8.6	8.6

TABLE 94

Dependence of Memory for Auditory Digits on Age (Ebbinghaus)
(Average Number of Errors per Pupil in Two Series)

AVERAGE AGE	8-DIGIT SERIES	9-DIGIT SERIES	10-DIGIT SERIES	6 TO 10 DIGITS
10.7	3.1	5.1	7.4	17.8
12.2	2.9	4.7	7.9	17.5
13.2	1.5	2.6	4.2	9.1
14.4	1.6	3.0	4.9	10.5
15.5	1.0	2.1	3.7	7.6
17.1	0.8	1.4	3.9	6.5
18.0	0.9	1.4	3.4	6.1

TABLE 95

Memory for Letter Squares, in Relation to Age and Practise (Winch)

SCHOOL GRADE	NUMBER TESTED	AVERAGE AGE	AVERAGE SCORE			Average for 3 Sets
			1st 10 Tests	2d 10 Tests	3d 10 Tests	
Ex-vii.	5	14 yrs. 3 mos.	23.8	29.0	31.7	28.1
vii.	5	13 " 5 "	26.3	27.9	31.1	28.4
vi.	5	12 " 3 "	26.8	32.0	34.6	31.1*
v.	5	11 " 4 "	18.4	22.9	26.3	22.5
iv.	6	10 " 5 "	21.3	24.8	26.6	24.2
iii.	6	9 " 0 "	14.1	17.7	19.7	17.1
ii.	6	8 " 2 "	13.2	16.8	17.2	15.7

*The girls of this group proved to have special ability.

TABLE 96

Memory for Letter Squares, Score for 10 Trials (Anderson)

AGE	CASES	MEAN	MINIMUM	MAXIMUM
8	52	108.2	47	198
9	92	109.7	36	182
10	115	127.7	35	213
11	126	139.8	60	264
12	139	157.8	76	272
13	125	156.9	52	298
14	96	165.6	74	283
15	58	170.8	67	323
16	25	181.6	104	318

TABLE 97

Memory for Concrete Words in Relation to Age and Sex (Pyle)

SEX	AGE	8	9	10	11	12	13	14	15	16	17	18	ADULTS
Male	Cases	34	58	64	55	60	60	35	25	14	7	5	64
	Mean	31.2	32.4	35.8	37.7	37.7	38.3	40.0	40.2	43.4	45.7	49.0	44.3
Female	Cases	37	68	69	52	70	51	34	13	17	8	2	88
	Mean	32.9	32.7	39.6	37.7	38.7	40.4	44.2	42.0	42.5	40.5	52.0	47.6

TABLE 98

Memory for Abstract Words in Relation to Age and Sex (Pyle)

SEX	AGE	8	9	10	11	12	13	14	15	16	17	18	ADULTS
Male	Cases	34	58	63	55	60	60	35	25	14	7	5	62
	Mean	22.9	26.3	26.8	31.7	31.0	32.4	37.3	34.1	40.0	41.1	40.8	42.3
Female	Cases	37?	68	69	52	69	52	31	13	17	9	2	88
	Mean	20.5	24.0	31.0	31.8	34.0	36.0	39.0	37.8	41.0	37.0	49.0	39.8

TABLE 99

Average Number of Sentences Correctly Repeated (After Squire and Carpenter)

AGE	6	7	8	9	10	11	12	13	14
Squire	7.8	8.4	9.8	10.1	10.9	10.9	13.5	14.5	--
Carpenter	5.3	6.0	6.2	7.2	7.4	7.7	8.0	8.3	8.5

TABLE 100

Memory for Pictures of Objects (After Squire and Carpenter)

AGE	6	7	8	9	10	11	12	13	14
Squire	5.3	6.5	9.5	9.8	9.1	11.4	10.5	10.0	---
Carpenter	5.3	6.1	6.0	7.3	7.5	7.5	8.4	9.2	10.0

In comparing new data with these results, differences in method must always be kept in mind. The differences between Smedley's conduct of the memory-span test and that prescribed above has already been described. The data of Table 92 are shown graphically in Fig. 71. It will be noted that Table 93 deals with averages, Table 91 with standards of performance, Table 94 with number of errors. Supplementing Tables 91 and 93, W. V. Bingham has reported to me the following results for auditory memory span for digits, secured from some 200 Dartmouth freshmen: median 7, P. E. 0.34, range from 5 or less to 12.

I am not able to explain the differences between the averages for letter-squares reported by Winch and by Anderson, save on the basis of selection of *S*'s: Anderson's results were obtained under my direction at Ithaca, N. Y., and include children in the public schools with no attempt at selection.

With reference to Table 99, the discrepancies between the results of the investigators are due primarily to two causes: Mrs. Squires' data refer to unretarded children only, and are based, as already explained, upon a different set of sentences. That the first of these factors is the more important seems indicated by similar divergencies in Table 100, where the experimental conditions must have been nearly identical. Other norms for sentences based upon results with the Binet tests include as standards: capacity to repeat a sentence of 6 syllables at 3 years, of 10 syllables at 5 years, of 16 syllables at 6 years and of 26 syllables at 12 years. It is unnecessary to add that much depends upon the sentences.

With regard to Table 100, it must be remembered that the averages given are based upon a presentation of a group of 30 objects: the use of 13 objects as specified in the directions above will yield somewhat smaller averages; competent adults usually get but 11 objects after an exposure of 15 sec. For the exposure of 13 pictures Binet reported the following average performances: at 7 years, 4.3 pictures; at 9 years, 6.2 pictures; at 11 years, 7.2 pictures.

(2) *Dependence on age.* That memory capacity increases in general from the early to the late school years is illustrated in Tables 91 to 101. The general evidence is fairly clear that this improvement is steady up to puberty, but that it suffers fluctuations after that period (see Tables 92, 94, 96, 97, 98). Several investigators adduce evidence that corroborates the popular notion that there exists a special 'memory period,' or stage of maximal efficiency somewhere in the 'teens,' when memory is stronger than it is later. For example, the very careful work of Pohlmann, with varied materials and varied forms of presentation, yields the net results (method of retained members) shown in Table 101, in which maximal efficiency is indicated at 14, followed by fluctuations, without real improvement through

the adolescent period. Bourdon could discern progress from 8 to 13, but not from 14 to 20. Bernstein and Bogdanoff, in testing memory for geometrical figures by the method of recogni-

TABLE 101

Net Efficiency of Various Memories, in Relation to Age (Pohlmann)

AGE	9	10	11	12	13	14	15	16	17	18	19	20
Average Capacity	39.4	41.4	55.7	59.1	62.1	68.9	55.3	62.9	58.6	58.0	65.4	68.3

tion, found that 23 *S*'s aged 14 to 15 averaged better than the 55 adults that they tested. Wessely, who tested retention during a long period (1 and 2 years), was convinced that ability to retain and reproduce poems is maximal at the years 12 to 14, and that vocabularies (Latin-German) are reproduced more accurately at the expiration of 1 to 4 weeks, when learned by 12-year-old, than when learned by 15-year-old *S*'s. Similar assertions concerning the relative amount of retentive capacity for poems by children and by adults are made by Larguier (48, 185 ff.), while Binet (8, 259 ff.) believes that children have the better retentive capacity, and adults the better attentive capacity.

Over against this evidence for a decline of efficiency after 14 we have the figures of Jacobs (Table 93) and the emphatic statement of Smedley (70, p. 49), based upon his Chicago results (Table 92), that "there is no 'memory period,' no period in early school life when the memory is stronger than it is at any later portion of the child's life." Smedley's records do, indeed, show that "auditory memory develops rapidly up to about 14 years of age, and but slowly after this period. The visual memory seems to develop rapidly up to about 15 or 16 years of age." . . . "It will be noted [Fig. 71] that, in the early life of the child, the auditory memory is stronger than the visual memory; after about 9 years of age, the visual memory of most of the children becomes stronger than the auditory

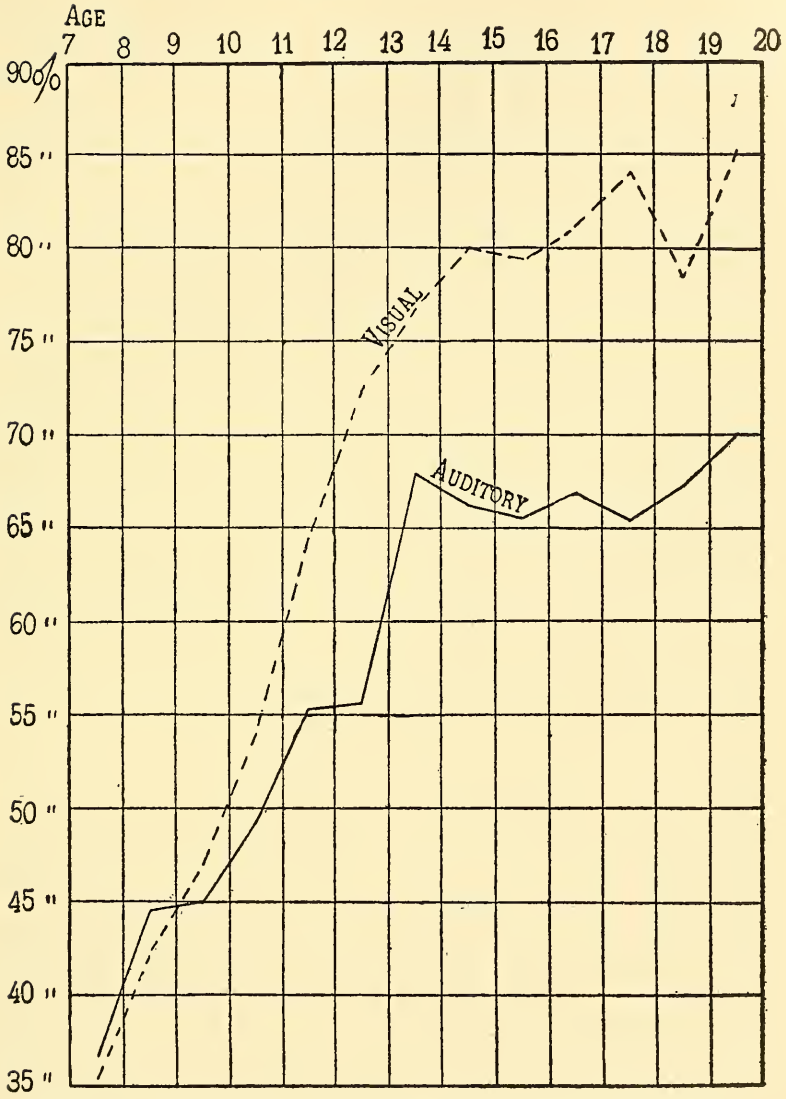

FIG. 71. DEVELOPMENT OF MEMORY FOR DIGITS (From Smedley).

memory, and continues to develop more rapidly than the auditory memory throughout school life. Yet, even in the high school, there still remains a small proportion of the pupils whose hearing memory is the stronger."

The dependence of different types of memory upon age has been studied especially by Netschajeff and by Lobsien. They agree substantially that, while the various forms of memory improve with age on the whole, there are periods of rapid development, followed by no improvement or even by a reduction; that while, on the whole, the greatest improvement occurs during the years 10 to 12, and development is retarded after 14, yet the different forms of memory, considered specifically, develop at different rates, and at periods that may not coincide in the two sexes. Thus, in boys, memory for objects is at first best developed, then follow, in order of chronological development, memory for visual terms, for acoustic terms, for actual sounds, for tactual terms, for numbers, for abstract terms, and finally for emotional terms. For girls, the chronological order is: visual terms, objects, sounds, numbers, abstract terms, acoustic terms, tactual terms, emotional terms. Special stress is laid upon the parallelism of development between memory for numbers and memory for abstract terms.

In Meumann's word-list tests, those types of error that indicate poor intelligence decreased with age, until, at 14 and 15, instances of misunderstood abstract terms were limited to about 10 per cent. of his *S*'s, while meaningless fusions, meaningless insertions, and the substitution of concrete for abstract terms had nearly disappeared, and the memory for abstract terms had so increased as frequently to be superior to that for concrete terms. It follows that age must always be taken into account in the interpretation of this test, particularly in estimating intelligence by it.

Since in a memory test so much depends upon the conditions of presentation, as will appear in what follows, I am inclined to regard many of these generalizations as of significance only under the particular conditions of the testing. The one safe generalization as to dependence on age would appear to be that made at the outset, viz.: capacity for immediate verbatim re-

production increases decidedly from early life to puberty, particularly during the period between 10 and 12, and increases more slowly and with fluctuations from puberty to maturity. The results of Cohn and Dieffenbacher, not here reproduced in detail, also accord entirely with this conclusion, as do the position of the medians in Anderson's percentile curves (Figs. 72 and 73).

(3) *Dependence on sex.* In general, girls pretty certainly surpass boys in immediate memory, but the differences are not always marked and perhaps do not extend to all forms of material. Investigations that agree in showing a general superiority of girls and women over boys and men are those of Anderson, Burt, Bolton, Calkins, Kirkpatrick, Pohlmann and Schuyten. Burt and Moore state that only 12.6 per cent. of boys exceed the median of girls, and add that "feminine superiority is a constant phenomenon in memory tests of every kind. It matters little what the age or training of the subjects may be. Hence, it is one of the best attested sex-differences and one of the most likely to be innate." [14]

The results of Anderson's extensive tests with letter squares have been summarized in Table 96. From Anderson's original data there have been arranged, after a preliminary process of numerical 'smoothing,' the percentile curves shown in Figs. 72 and 73.

Curves of this construction are so valuable for diagnosing the station of any individual's performance that these curves are here reproduced, despite the fact that they exhibit a number of irregularities due to the small number of cases available at certain ages. I have also calculated from Anderson's data the tables of distribution (102 and 103). It is therefore possible for any *E* to use these data by adding to them further scores and then recasting the percentile curves to accord with the combined data. For this purpose the scores may be considered as if they were the middle points within the ranges here indicated; for example, the distribution for 9-year-old boys may be read, 3 scores of 65, 4 scores of 75, etc.

By inspection of the medians (50th percentiles) in these charts it will be seen that girls are inferior to boys at 9 years

[14] I find, however, some difficulty in identifying the figures upon which this conclusion is based, as they appear in Burt's different articles not to refer to the same groups.

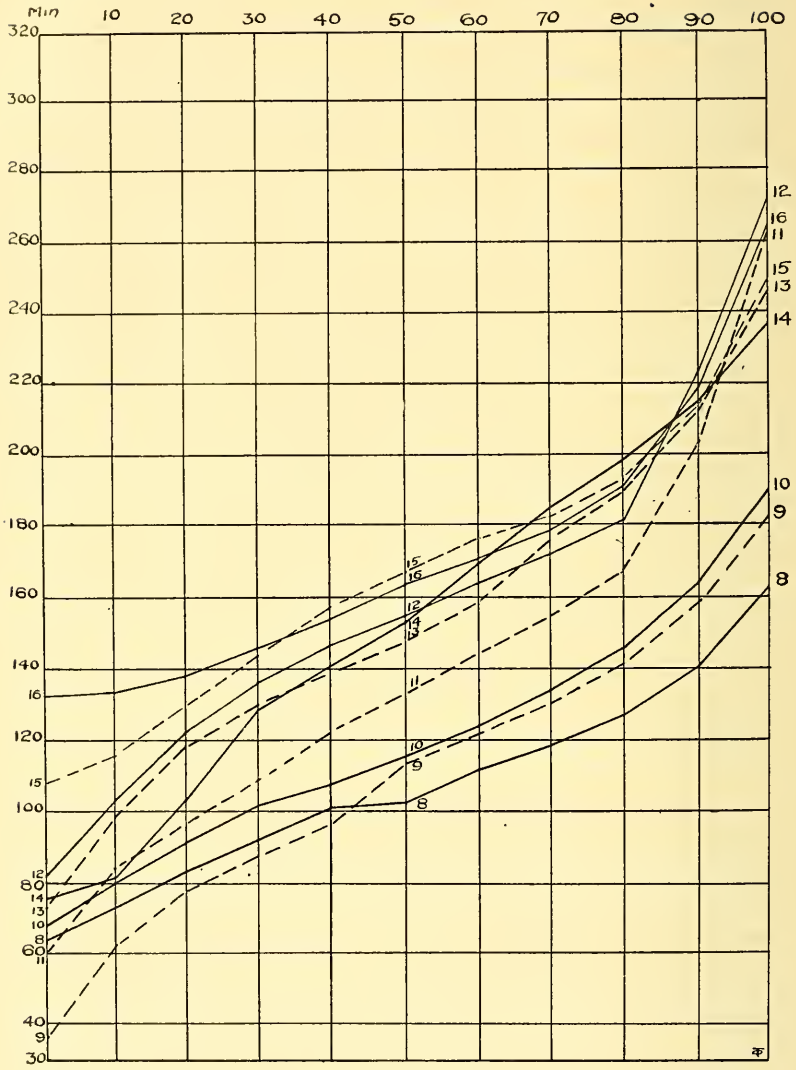

FIG. 72. PERCENTILES OF MEMORY FOR LETTER SQUARES, BOYS (Anderson).

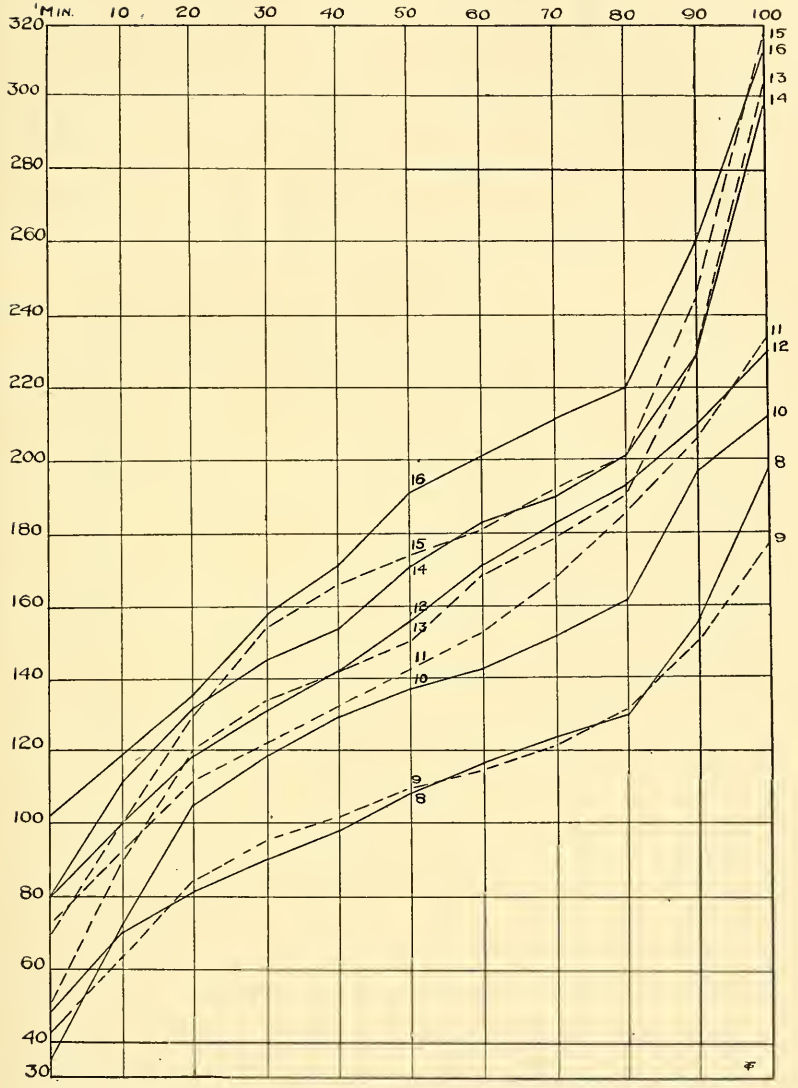

FIG. 73. PERCENTILES OF MEMORY FOR LETTER SQUARES, GIRLS (Anderson).

and practically the same as boys at 12 years, but that elsewhere the girls are superior and that their superiority is especially striking in the higher percentiles. The curves are undoubtedly affected by a poor group of girls at 9 years and an unusually good group of boys at 12 years, since the progress with age is decidedly broken at these points in the manner mentioned. On the whole, Anderson figures, girls are some 27 per cent. better than the boys in letter squares.

TABLE 102

Distribution of Boys' Scores in Letter Squares (After Anderson)

AGE	8	9	10	11	12	13	14	15	16
60- 69	1	3	1	1					
70- 79	1	4	4	3		1	2		
80- 89	2	5	2	3	3	2	1		
90- 99	5	5	7	9	2	2	2		
100-109	5	3	10	2	3	6	1	2	
110-119	0	4	9	5	3	3	1	2	
120-129	5	6	4	4	4	3	7	1	
130-139	1	4	3	5	9	9	2	3	3
140-149	0	3	6	9	6	9	4	3	1
150-159	0	4	3	5	11	8	3	2	1
160-169	1	1	3	4	7	4	3	2	1
170-179		0	1	4	9	3	2	5	2
180-189		1	0	3	3	4	3	5	1
190-199			1	2	1	5	4	2	1
200-209				0	2	2	3	2	1
210-219				0	3	1	3	0	0
220-229				0	1	3	1	1	0
230-239				0	2	2	1	0	0
240-249				1	1	1		0	0
250-259				0	0			1	0
260-269				1	1				1
270-279					1				
Total--	21	43	54	61	72	68	43	31	12

In other investigations the superiority of girls is either less clearly evident or exhibited in some aspects of the tests only. Thus Lobsien's tests with varied materials (Table 106) likewise showed that girls reproduced more, but that boys were more

TABLE 103

Distribution of Girls' Scores in Letter Squares (After Anderson)

AGE	8	9	10	11	12	13	14	15	16
30- 39			1						
40- 49	1	2	0						
50- 59	0	0	0			1			
60- 69	1	1	1			1		1	
70- 79	2	3	2	2	1	0	1	0	
80- 89	5	4	3	1	1	2	2	0	
90- 99	4	5	1	3	2	2	1	1	
100-109	4	9	4	5	5	2	0	1	1
110-119	2	11	4	3	4	0	1	0	0
120-129	6	2	7	10	4	5	2	0	1
130-139	3	4	10	6	7	6	5	2	0
140-149	1	5	10	7	8	9	6	0	1
150-159	0	1	2	6	5	3	3	4	1
160-169	0	1	4	6	3	3	6	3	1
170-179	0		2	3	7	5	5	2	2
180-189	1		0	2	4	6	6	1	0
190-199			1	4	6	3	3	7	1
200-209			1	2	4	0	3	0	1
210-219				1	3	0	3	0	3
220-229				1		2	1	0	0
230-239						0	1	0	0
240-249						0	0	0	0
250-259						1	0	0	0
260-269						1	0	0	0
270-279						1	0	0	0
280-289						0	1	0	0
290-299						1		0	0
300-309								0	0
310-319								0	1
320-329								1	
Total..	31	48	61	65	67	57	53	26	13

apt to get the o..er right.[15] Netschajeff also concluded that girls made more illusory errors (especially at ages 9 to 11). He also found that boys had the better memory for real objects, girls for numbers and words, in which they surpassed boys, particularly during the years 11 to 14. Wissler's tabulation

[15]Note analogous results in the Test of Report (No. 32).

of the freshmen tests at Columbia University and Barnard College reveals sex differences in memory span for digits that are less than the P.E. of the averages, and that favor the men for auditory, and the women for visual series (Table 105).

TABLE 104

Percentage of Accuracy in Memory for 2-place Numbers (Schuyten)

		MORNING	AFTERNOON
First test (Afternoon first)	Boys	58.1	64.0
	Girls	69.6	77.5
Second test (Morning first)	Boys	57.9	35.0
	Girls	62.6	55.1

TABLE 105

Sex Differences in Memory Span for Digits in College Freshmen (Wissler)

	AUDITORY PRESENTATION			VISUAL PRESENTATION		
	Number	Average	P. E.	Number	Average	P. E.
Men	266	7.6	0.4	142	6.9	0.5
Women	42	7.3	0.5	42	7.3	0.4

Pyle's averages for memory for words show possibly an advantage for the girls, but the differences are only slight. Cohn and Dieffenbacher, similarly, find that girls surpass boys only when groups of the same school grade and same social status are compared, and that when the results are plotted by age the curves cross each other six times, so that the sex-difference which appears in lump comparisons turns out to be practically an accident.

(4) *Dependence on practise.* (*a*) *General.* Practise produces a measurable increase in the memory span (Bolton). In the use of nonsense syllables, indeed, the practise effect can be dis-

cerned even at the expiration of 60 days of experimental work (Müller and Schumann).

Winch, from his use of the letter-square, as well as of auditory letter series, not only declares that there is a "marked and almost invariable improvement," but "that 'pure memory' is markedly improvable by practise" (80, p. 134). Thus, 38 S's, ages 8 to over 14, obtained, in 3 sets of 10 tests each (1 week between the 1st and 2d, and 2 weeks between the 2d and 3d), the average scores 20.6, 24.4, and 26.6 (averages of the scores of Table 95).

(*b*) The *transfer of practise* from the specially trained form of memory to other forms of memory would appear, from theoretical grounds, to be limited to those cases in which the material, content, or method of procedure of the other forms were related to the material, content, or method of procedure of the trained form. This is essentially the conclusion reached by Ebert and Meumann (25, p. 200), who say: "The objective results of our experiment show that special memory practise is accompanied by a general improvement of memory. This concomitant improvement does not, however, extend equally to the other 'memories,' but appears to follow the law that the specific memories participate in the improvement directly in proportion as they are related in content, or in media and method of learning to the specific memory that was trained."

Winch has been led, by experiments in memorization of poetry and historical prose (81a), to take the more radical stand that "improvement, gained in practise in memorizing one subject of instruction, is transferred to memory work in other subjects whose nature is certainly diverse from that in which the improvement was gained, . . . at least so far as children of these ages and attainments are concerned." Again, in his second paper (81b) he concludes that "improvement through practise in rote memory for things with and without meaning is followed by improvement in substance memory for stories," and this even though the correlation between the two functions is very low and even doubtful.

On the other hand, Starch found no improvement in auditory

memory span after 14 days of drill in mental multiplication, and Sleight found no general improvement in his drilled sections, save that *S*'s drilled in memorizing poetry or tables showed subsequent improvement in memory for nonsense syllables, on account, he thinks, of the use of rhythm in these types of memorizing. Drill in memory for 'prose substance' improved that sort of memory, but no other, and even worked disastrously for subsequent memorizing of nonsense syllables. The conclusions reached by Ebert and Meumann have been criticized by several writers and directly controverted by Wessely, who says that for memory there seems to be no formal practise effect.[16]

(5) *Dependence on fatigue.* Though fatigue may affect immediate memory and undoubtedly does so when severe, it is difficult to arrange memory tests that will serve as useful indexes of fatigue, particularly because either practise or ennui affects the results more than does true fatigue. On this point we find Bolton, Ebbinghaus, Schuyten and Smedley in agreement. Two investigators, however, have secured results worth mentioning. Winch (82) divided pupils into two equivalent sections and practised them with letter-squares until the rank-orders were 'steady.' He then continued the tests with one group in the morning, with the other in the afternoon. Both groups showed improvement, but the morning workers improved from 2 to 6 per cent. more than the afternoon workers. Ritter gave up the determination of fatigue by span tests with numerals, but he did achieve results which he considers of special value by the use of 6-term series of two-syllabled nouns. With this material, he finds that errors increase with fatigue, and he goes so far as to assert that this test is the best one available for the investigation of fatigue.

---

[16]Allusion may be made in this connection to the evidence for transfer found by Dallenbach with experiments in the analogous field of visual apprehension (Test 25). Aside from the fact that Sleight worked with pupils as well as adults, it is possible that the discrepancies noted here concerning the transfer of practise may be due to the ages of the *S*'s studied. I have suggested elsewhere (*JEdPs*, 5: 1914, 362) that, particularly in the case of transfer, experimentation with children has been neglected and that results secured with adults may not necessarily apply to the mental processes of children.

(6) *Dependence on physical capacity.* Both Netschajeff and Smedley find that pupils that are larger, stronger, and better developed physically have better memories than those of the contrary type. "This suggests," says Smedley, "that the immediate sense memory is dependent upon good brain formation and nutrition." (See 70, pp. 58-59, for numerical evidence.)

Miss Kohnky used memory span for digits among other tests in an investigation of the effect of dental treatment upon pupils in a 5th-grade Cincinnati school. The series with 6 digits proved too easy for this work; the series with 8 digits yielded virtually the same results for treated and untreated pupils, but the series with 7 digits yielded a gain of 10.8 per cent. between tests made in October and in the following May in the case of the pupils having dental treatment, as contrasted with no gain in the pupils without treatment.

(7) *Dependence on the nature of the material.* (*a*) When *digits and consonants* are given under the same conditions, digits are easier to reproduce (Jacobs, Sharp), especially during the years 8 to 13 (Bourdon). But, if 10-place series are presented auditorily, thrice, the order of excellence for recall is (1) consonants, (2) names of objects, (3) 2-place numbers, (4) nonsense syllables (Pohlmann).

(*b*) Netschajeff, Lobsien, Pohlmann, and less elaborately Kirkpatrick and Calkins, have compared memory for series made up of real objects, of numbers, of sounds, and of words

TABLE 106

Memory for 9-term Series of Different Kinds (Lobsien)

KIND OF SERIES	SCORE IN PER CENT. CORRECT	
	Boys	Girls
Real objects	82.2	91.4
Auditory numbers	64.8	71.8
Sounds	59.6	62.2
Tactual terms	64.2	71.0
Visual terms	60.6	67.2
Auditory terms	59.4	60.2
Emotional terms	31.2	59.4
Foreign terms	24.0	23.8

having characteristically visual, auditory, tactual, or emotional associative meanings. Table 106 gives illustrative results from Lobsien. Pohlmann, however, concluded that the assumption of Netschajeff and Lobsien that the presentation of visual, auditory, and other terms arouses the visual, auditory, and other imagery that their meaning implies, is erroneous, so that the results of these investigations are of little real significance.

Kirkpatrick, and after him Miss Calkins, found, like Netschajeff and Lobsien, that memory for objects (or pictures of objects) was superior to that for words, both for immediate and for delayed reproduction; in the latter, for example, there were recalled seven times as many objects as words. The same investigators determined the order of excellence for recall of different kinds of words to be:—visual terms, auditory terms, names of objects.

(*c*) Up to the 12th year, *concrete words* are reproduced better than abstract words, but 14 and 15-year-old *S*'s frequently make better records with the latter, according to Meumann; but according to Pyle, concrete words are reproduced better at every age. The average difference amounts in Pyle and in

TABLE 107

Memory for Related and for Unrelated Words (Norsworthy)

Age	Related Words, 288 Cases				Unrelated Words, 270 Cases			
	boys		girls		boys		girls	
	Median	P. E.	Median	P. E.	Median	P. E.	Median	P. E.
8	13.0	1.0	13.0	1.6	11.1	1.6	11.5	1.3
9	14.0	2.0	14.0	1.7	12.2	1.7	12.4	1.4
10	15.0	1.7	15.3	1.9	12.2	1.7	14.4	1.4
11	15.0	1.7	16.5	1.7	12.5	1.8	14.3	1.4
12	16.4	1.8	16.0	1.6	12.8	1.8	14.0	1.5
13	16.5	1.8	17.0	1.5	13.5	2.1	13.5	1.5
14	16.9	1.3	17.5	1.5	13.7	2.2	14.0	1.5
15	16.0	1.3	17.5	1.5	13.7	2.2	14.0	1.5
16	17.0	1.3	17.8	1.5	14.0	2.2	14.5	1.5
Adults	16.5	1.5	17.0	1.9	12.8	1.2	13.0	1.4

Burt (16) to 20 per cent., roughly, *i. e.*, most *S*'s recall about one-fifth more concrete than abstract words.

(*d*) *Related terms, i. e.,* a series of words not in a sentence, but readily associated with one another, are more easily recalled than unrelated words. For data, see Table 107 from Miss Norsworthy.

(*e*) Material so arranged as to aid *localization* is more easily remembered, especially by children. For example, 12 consonants in the letter-square form are easier to recall than 12 consonants in a single line; similarly, digits pronounced in rhythm are easier to recall than digits pronounced in even tempo (Müller and Schumann). Pohlmann found grouped series to be easier in 133 of 144 trials.

(8) *Dependence on sense-department directly stimulated.* It is evident that a complete isolation of the different modalities cannot be accomplished by different forms of presentation: *e. g.*, auditory-minded *S*'s may actually retain and reproduce impressions presented to the eye in auditory, or mainly in auditory terms, and so on (cf. Abbott and Finzi). It is also evident that what seem like minor variations in the manner of conducting the test may occasion considerable variations in the performance of *S*'s. These facts account for much of the divergence and seeming contradiction in the results of various investigators with regard to the relative advantage of addressing stimuli to different senses.[17]

With regard to the relative advantage of auditory over visual presentation Kemsies found presentation by ear the better for Latin words and for nonsense syllables; von Sybel found auditory presentation better than visual for both auditory and visual types of *S*'s; Henmon found as his most striking result a marked superiority of auditory over visual presentation for all of his *S*'s and for all forms of material. Hawkins reported that ten nouns heard are recalled better than ten nouns successively seen in the case of younger *S*'s, but that the reverse

[17]These divergencies have been well summarized by Henmon, to whose account the reader is referred for details of the conclusions reached by Meumann, Münsterberg and Bigham, Quantz, Lay, Itschner, Füchs and Haggenmüller, Cohn, Kemsies, Finzi, Fränkl, Segal, von Sybel, Schuyten, Pohlmann and others, together with his own conclusions.

holds true for above 15 years. Pohlmann's extensive experiments, which are criticized by Henmon because of being conducted by the group method, show that auditory presentation is better for meaningful material (words), while the reverse is true with non-significant material (digits and nonsense syllables). On the other hand, the superiority of visual over auditory presentation appears in the tables and charts of Smedley and in Chambers' results for 7th and 8th grade pupils.

With regard to the advantage of combined appeal to eye and ear or to eye, ear and motor memory (articulation or writing), there are similar discrepancies. The work of Pohlmann (Table 108) indicates a superiority of auditory-visual presentation over either auditory or visual presentation, alone—a result in accordance with Smedley's. Pohlmann also investigated the effect of these three forms of presentation upon numerals and nonsense syllables, with the result that for 230 *Volksschule* girls, using 10-term series, given thrice, the percentage of accuracy was, for visual-auditory 53 per cent., for visual 52 per cent., and for auditory 42 per cent., which agrees in substance, so far as it goes, with Smedley's results for digits. This investigator found the order of superiority to be: (1) auditory-visual-articulatory, (2) auditory-visual, (3) auditory-visual-hand-motor, (4 and 5) visual or auditory (depending on age). Illustrative figures for *S*'s aged 16 years are, for the five forms just

TABLE 108

Dependence of Memory upon Form of Presentation (Pohlmann)
(Percentage of Retained Members, 10-Term Series, 350 Pupils, 9-14 Years)

NATURE OF MATERIAL	FORM OF PRESENTATION	PERCENTAGE RETAINED
1. Actual objects	Shown and named by *E*	72⅛
2. Actual objects	Shown, only, successively	70
3. Names of objects	Seen and heard by *S*	
4. Names of objects	Heard, only, by *S*	55⅙
5. Names of objects	Seen, only, by *S*	50⅔
6. Names of objects	Seen, heard, and pronounced by *S*	49½

In the upper classes, 5 becomes superior to 4.

mentioned, 88.4, 86.9, 82.4 (circa), 80.0, and 66 per cent., respectively. Combined appeal is, then, most powerful, but the task of writing proves somewhat distracting. Münsterberg and Bigham conclude, similarly, that "a series of impressions offered to two senses at the same time is much more easily reproduced than if given only to sight or only to hearing." Thus, in the case of 10 numbers the per cent. of error for numbers heard, seen, and both seen and heard were 14.1, 10.5, and 3.9, respectively. On the other hand, Henmon concludes that visual-auditory-motor presentation is slightly inferior to the auditory and to the auditory-visual, but superior to the visual alone, and that visual-auditory presentation is slightly inferior to the auditory alone, while decidedly superior to the visual alone. In general, he found the advantage of combined presentation much less than that reported in earlier investigations. Kemsies discovered that visual-auditory presentation usually gave poorer results than visual or auditory alone in tests with Latin words and nonsense words.

The question as to whether articulation does or does not assist in subsequent recall is also answered differently by different investigators. Thus, Cohn found that in memorizing consonants all his *S*'s did best when they read aloud, less well when speech movements were suppressed, least well when numbers or vowels were pronounced as distractors during the reading of the consonants. Quite similar results which were reached by Lay in his investigation of the teaching of spelling have been contested by Itschner and by Fuchs and Haggenmüller. Henmon also declares that articulation or vocalization is of little value for immediate memory.

A closely related question concerns the possibility of determining *S*'s *ideational type* by scrutiny of his performance under different sorts of presentation. Fränkl and also Segal believe that visual presentation gives better results with visual types, auditory with auditory types, and Meumann concludes from such tests that in learning, better reliance can be placed upon *S*'s type than upon an appeal to several sense departments. But the evidence is fairly clear that, as Angell says, while memory tests "may certainly be so administered as to show over what

sensory arcs the best results may be achieved in assimilating information of various kinds . . . as objective tests of imagery apart from introspection, they have few virtues and no reliability."

(9) *Successive vs. simultaneous presentation.* If 15 words are exposed simultaneously or successively for equivalent lengths of time, successive presentation is easier for young, but simultaneous for older children, according to Hawkins.

(10) *Dependence on number of presentations: repetition.* Pohlmann, Lipmann, Smedley, and others have found that hearing a series thrice or twice, instead of once, improves its recall. However, Hawkins found two hearings less effective than one or three. It is certain that more is accomplished in the first hearing than in a large number of repetitions, and that the effect of repeated presentation is different in different *S*'s, so that individual differences are more marked after many hearings than after one hearing (Smith). Smedley's test of 38 10-year pupils, with auditory digits, gave, for the first hearing 47 per cent., for the second 55 per cent., and for the third 59 per cent. correct reproduction. In some of Smith's tests, 12 presentations did not double the efficiency attained in one presentation.

(11) *Dependence on rate and duration of exposure.* Bergström's tests indicate that nonsense syllables exposed at the rate of one in 0.77 sec., with durations of exposure of .041, .082, .164, and .318 sec. yield practically the same results, though there is a slight preference for .082 sec.

The same investigator found that, both with auditory letter and word series and with visual nonsense-syllables series, a relatively slow rate of exposure (1.5 to 2 sec. per term) yielded more accurate results than a faster rate (one term in a fraction of a second). The slower rate is especially helpful in lists of words, and for those *S*'s that try to develop associations between the terms as they are presented. Bergström summarizes by saying: "The acquisition and retention of a series of familiar associable words varies approximately as the logarithm of the interval at which the words are spoken" (6, p. 221).

(12) *Dependence on interval between presentation and repro-*

duction. Relatively short intervals make, apparently, but little change in reproduction. Thus, Winch could discern no clear differences in the reproduction by school children of letter-squares, with or without a 25-sec. empty interval between presentation and reproduction.

Binet and Miss Sharp compared immediate memory with 'recapitulatory' memory (memory of conservation); they both noted that the word lists in immediate reproduction seemed to be held largely by sound (so that, for example, such errors as *flower* for *floor* were common), whereas lists reproduced 3 min. later appear to be held more often by meaning, since "the errors are usually additional words suggested from analogy of sense" (*e. g., dog* suggested by *cat, cold* by *winter,* etc.).[19]

Attention has already been called (7, *b,* above) to the demonstration by Kirkpatrick and by Calkins that the reproducibility of different forms of material is not equally affected by a 3-day interval.

(13) *Effect of distraction.* Smith's use of the method of letter squares (71), with and without the distraction of concomitant activities, shows the order of efficiency under these conditions to be, from best to worst:—(1) without distraction, (2) with tapping to the beat of a metronome, (3) with repetition of a vowel, and (4) with counting by 2's or 3's. Cohn, with the same test, found that an auditory-motor *S* was more seriously disturbed by auditory-motor distractors than a visually minded *S*, and that, when such distraction is used, visual memory steps in to aid, provided *S*'s constitutional make-up (*Anlage*) will at all permit (22, p. 182).

(14) *Reliability.* With the exception of Brown, whose coefficients were only .50 to .68, investigators have found tests of immediate memory to yield a satisfactory degree of reliability. Examples are: Burt, .70 for one group of *S*'s, .93 for another; Wyatt, .75 for one group, .76 for another; Abelson, .74 to .81; Simpson, .73 for all *S*'s collectively.

---

[19]The tendency of adults is away from rote memorizing in favor of a memory of meanings. It would, then, be interesting to see whether children exhibited these same tendencies that Sharp's university students did, or exhibited them in as marked a degree.

(15) *Correlation with mental ability.* Bolton, Ebbinghaus (Table 109) and Wissler (who found a correlation of but 0.16 between class standing and the memory capacity of 121 Columbia freshmen) seem to be the only investigators to deny a relationship between immediate memory and intelligence. Jacobs, at the other extreme, asserts that there is a "notable concomitance" between school standing and "span of prehension." The truth would appear to lie, as usual, between these extremes. The more careful correlational work of the past few years demonstrates at least a fairly good degree of correspondence between immediate memory and either school standing or estimated general intelligence. The several investigators who have found correspondence of this sort have expressed their conclusions with certain restrictions or explanations, as will appear in what follows. There is some evidence, for example, to indicate that the correspondence between immediate memory and school standing is closer in the lower than in the higher grades, and that the correspondence with general intelligence is closer with tests of delayed than of immediate recall.

TABLE 109

Relation of Memory for Auditory Digits and Intelligence (Ebbinghaus)

	AVERAGE NUMBER OF ERRORS PER PUPIL		
	Bright Group	Average Group	Dull Group
9-digit test	84	87	84
10-digit test	147	147	135
6-10 digit tests, collectively	318	319	303

With backward children Abelson found rather low correlations, 0.18 and 0.19, between memory for names and imputed practical intelligence (competence to perform errands), but higher correlations with school performances (0.20 and 0.24 with estimated ability in reading and 0.30 and 0.32 with estimated ability in arithmetic).

Binet (8) contrasted 6 dull and 5 bright boys, and found that, on the whole, the latter surpassed the former in memory: the difference, as in his tests of other traits, was, however, more evident at the first, than at any subsequent trial.

Brown found correlations of 0.40 to 0.59 with school marks and 0.49 to 0.55 with estimated general intelligence.

Burt estimated intelligence in various ways and measured memory for concrete words, abstract words and nonsense syllables. The corrected correlations for estimated intelligence and general standing in the memory tests were 0.60 for Elementary School, and 0.82 for Preparatory School boys. The uncorrected correlations were higher for memory for concrete words (0.58 and 0.84) than for memory for abstract words (0.48 and 0.78); those for nonsense syllables were 0.43 and 0.75, while the amalgamated memory tests correlated by 0.67 and 0.69 with examinations in mathematics and by 0.82 with examinations in literary subjects (in the Preparatory School). Burt argues from these figures that the current examination system stresses ability to remember. Meumann's conclusion that bright children display a relatively superior ability in memory for abstract terms was not substantiated by Burt.

Cohn and Dieffenbacher divided their *S*'s into two groups—the better and the poorer intellectually. The former excelled the latter in the memory tests in 11 of his 14 groups; the superiority amounted to about 10 per cent., and turned out to be less in the higher than in the lower school grades.

Lapie contrasted pedagogically advanced with pedagogically retarded pupils, and concluded that these groups differed little in retentive power as such, but that the retarded pupils frequently reproduced the material in bizarre and contradictory combinations, *e. g.*, as in speaking of "a young peasant 54 years old."

Meumann says that the quantity of material reproduced is not in itself a reliable index of intelligence, yet the average results of mass experiments will always show that the more intelligent *S*'s have the better memory efficiency. His own experiments, he declares, were so extensive and so carefully executed as to leave no doubt at all upon this point (51, p. 78). More reliable, however, are the qualitative results attained from memory tests of the form used in his own experiments. Here, he says, virtually complete coincidence is found between the several indexes of intelligence, and between them and the school marks and the estimate of mental ability by teachers. Certain characteristic indexes of poor intelligence, however, such as the fusion of abstract terms into meaningless collocations, may not be shown by all of the stupid children; if they are shown, they form a reliable index of poor intelligence, while if frequent, they indicate not only poor intelligence, but also the lack of moral qualities, such as self-control and carefulness. Incidentally, Meumann points out that, in theory, we should distinguish carefully between natural ability and actual ability as shown in school performance; these, nevertheless, tend to coincide in practise.

Pohlmann, like Binet, dealt with contrasted groups. He concludes that, while in general the better pupils have better memories, there are numerous exceptions, particularly in that poor pupils may do as well as bright pupils in the memory tests.

The two contrasted groups of adults tested by Simpson were fairly well separated by his tests of memory for words; none of the poor group reached the median performance of the good group, and only 10 per cent. of the poor group were as good as the lowest 6 per cent. of the good group. The correlation with the estimated intelligence of the good group was 0.93.

Smedley declares that the "parallelism between school standing and memory power holds good throughout school life" (70, p. 54), and demonstrates this by reference to mass results distributed to show the memory capacity of pupils of a given age in different grades (Fig. 74), or the

capacity of pupils at and above grade as compared with the capacity of pupils below grade at different ages (Table 110).

Winch's letter-square tests convince him that "general mental ability [rank in examinations in reading, arithmetic, dictation, and English composition] is accompanied by 'good memory.'" "With two exceptions, no girl whose memory mark is relatively low has a high place in class." "'Good memory,' though usually accompanied by general efficiency, is not invariably so." Again, Winch contrasted six 13-year-old girls, who stood between Number 1 and Number 11 in a class of 35, with 6 girls of the same age, who stood 25th to 30th in a class of 30, and found the average score of the bright girls to be 26.9, as compared with an average score of 19 for the dull girls (80, p. 133).

Wessely believes that the correlation between memory and class standing is more evident in lower than in higher grades—a view expressed also by Cohn and Dieffenbacher and which, if confirmed, might be explicable by the tendency to put a premium upon memorization in the lower grades.

Wyatt's tests with nonsense syllables gave as a correlation with intelligence 0.59, P.E. .07, for immediate and 0.74, P.E. .06, for delayed (2 days) reproduction; his tests with letter squares gave a negligible correlation 0.18, P.E. .11.

(16) *Memory of defectives.* Galton applied Jacobs' tests to imbeciles, and found that most *S*'s of this type failed to repeat more than 4 digits, while several imbeciles who had remarkable memories for dates or for passages in books showed complete failure (span not over 3) in memory for digits. Johnson computes the average span for feeble-minded (selected *S*'s of the so-called 'school-case' group) at 5.3, or approximately 1.3 digits less than the normal span of an 8-year child. The distribution of efficiency, as he found it, is shown in Table 111. Johnson comments upon the fact that the difference between the memory

TABLE 110

Relation of Memory for Digits and School Standing (Smedley)

AGE	NUMBER TESTED	AUDITORY		VISUAL	
		Average Standing of Pupils At and Above Grade	Average Standing of Pupils Below Grade	Average Standing of Pupils At and Above Grade	Average Standing of Pupils Below Grade
9	99	47.8	39.7	50.3	41.9
10	88	54.4	42.7	61.6	46.2
11	91	59.0	48.6	69.4	53.3
12	92	62.6	52.2	76.7	66.0
13	110	70.4	64.3	80.7	72.3
14	116	68.9	62.6	87.6	74.9
15	94	68.9	62.4	80.9	75.0
16	75	70.1	65.8	83.3	78.8
17	56	67.5	62.7	87.8	81.2

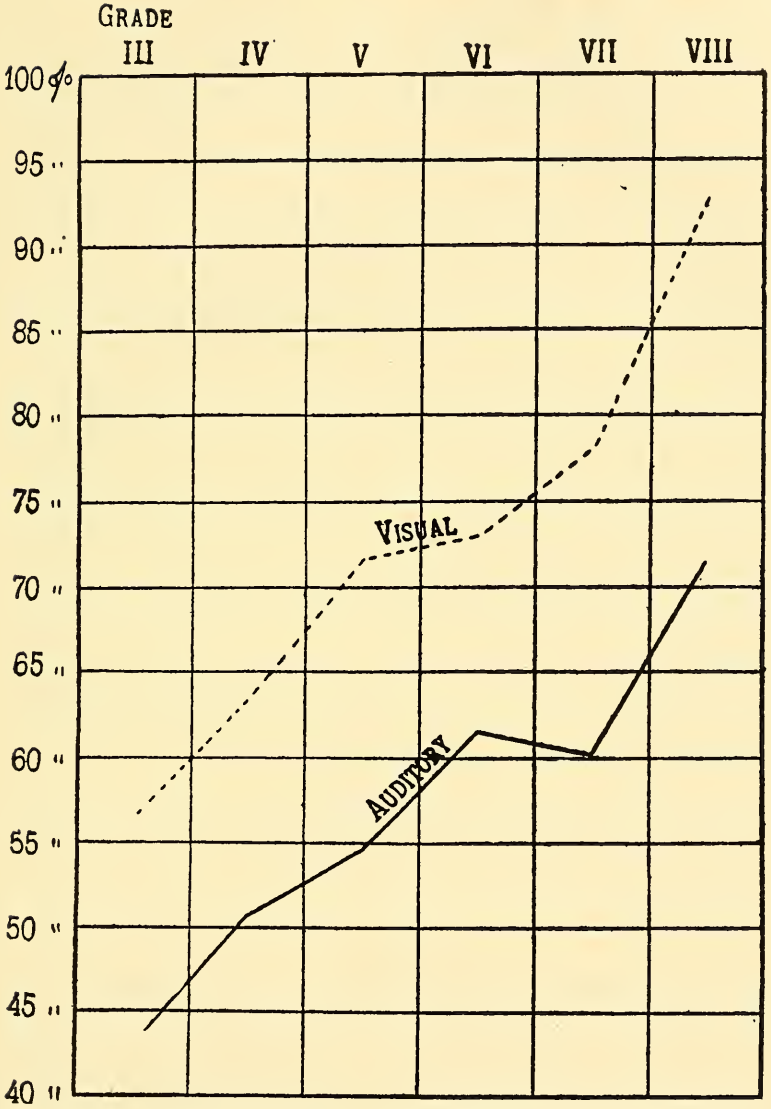

FIG. 74. MEMORY CAPACITY OF 12-YEAR-OLD PUPILS, BY GRADES (From Smedley).

TABLE 111

Memory Span for Digits in the Feeble-Minded (Johnson)

Number of digits	3	4	5	6	7	8
Repeated correctly by	70	66	51	27	14	4

NOTE—The larger groups include the smaller ones at their right.

span of the feeble-minded and of normal children seems to be of a smaller order than the general difference in intellectual ability of the two groups.

Miss Norsworthy compared normal and feeble-minded children with respect to memory for related and for unrelated words. Her standards for normal children have already been reported (Table 107): the relation of feeble-minded to normal efficiency is shown in Table 112. The figures are to be interpreted simply: five per cent. of the feeble-minded do as well with the related-word test as do 50 per cent. of normal children, etc.

Smedley states "that the boys of the John Worthy School [incorrigibles, defectives, truants, etc.] are lower in memory power than are the pupils of the other schools, and this disparity increases with age" (70, p. 59).

Smith's tests with epileptics (73) show that, in the auditory letter-span test, they are generally inferior to normal *S*'s, and in particular, that they make nearly three times as many errors of insertion.

TABLE 112

Comparative Memory Capacity of Normal and Feeble-Minded Children (Norsworthy)

	ABOVE MEDIAN	ABOVE −1 P. E.	ABOVE −2 P. E.
Normal (both tests)	50	75	91
Feeble-minded, in related words	5	19	30
Feeble-minded, in unrelated words	6	18	27

(17) *Other correlations.* Krueger and Spearman found no correlation between memory for digits (serial visual exposure) and either ability to add, to discriminate pitch, or to discriminate dual cutaneous impressions.

Memory for digits and memory for letters were correlated to a high degree in Miss Sharp's *S*'s, while memory for short sentences correlated best with memory for letters.

Smedley studied the relation of memory for digits and ability to spell, and concluded that "while, on the whole, the good spellers have decidedly better memory power than the bad spellers, yet there are individuals among the poor spellers who are superior in memory power, and individuals among the best spellers whose memory power is scarcely up to the average of their age. While this native power of sense-memory plays an important rôle, it is by no means the only factor in learning to spell" (70, p. 61).

Abelson found the following correlations with memory for words in his study of 88 backward boys and 43 backward girls: interpretation of pictures, boys 0.30, girls 0.33; memory for sentences, boys 0.66, girls 0.42; tapping, boys —0.08, girls 0.30; memory for commissions, boys 0.38, girls 0.34.

Brown tested several groups, mainly pupils 11 to 12 years old, and found the following correlations with his test of learning nonsense syllables: completion test, 0.28, 0.37, 0.52; memory for poetry, 0.38, 0.49; speed in addition, —0.13, 0, 0.27; accuracy in addition, —0.23, 0, 0.31; drawing, 0.39.

Burt's correlations, so far as they apply to tests mentioned in this work, are displayed in Table 113, Wyatt's in Table 114.

Simpson publishes the following as estimated true correlations with memory for words for people in general: completion test, 0.82; hard opposites, 0.84; easy opposites, 0.65; the a-test, 0.54; memory for passages, 0.80; adding, 0.39.

In the half dozen rather varied tests classed together by Heymans and Brugmans the intercorrelations for the several tests ranged from —0.34 to +0.71. The pooling together of the results of tests classed as tests of memory, imagination, etc., yielded correlations of 0.75 between memory and imagination,

TABLE 113

Correlations with Amalgamated Results of Immediate Memory Tests (After Burt)

RELATED TEST	RAW CORRELATIONS		CORRECTED CORRELATIONS	
	Elem. Sch.	Prep. Sch.	Elem. Sch.	Prep. Sch.
Spot pattern	.25	.55	.41	.84
Mirror drawing	.08	.44	.13	.64
Tapping	.01	.52	.01	.80
Pitch discrim.	.13	.20	.19	.27
Lifted weights	.05	.15	.07	.22

TABLE 114

Correlations Between Memory Tests and Other Tests (After Wyatt)

	DELAYED MEMORY	IMMEDIATE MEMORY	LETTER SQUARES
Analogies	.70	.64	.28
Completion	.72	.61	.03
Word-building	.66	.57	.00
Part-wholes	.51	.52	.09
Fables	.43	.41	.31
Immediate mem.	.71	--	.25
Letter squares	.15	.25	--

TABLE 115

Recall of Different Members of a 7-Term Series (Binet and Henri)

Place in series	1	2	3	4	5	6	7
Times recalled correctly	143	139	115	111	122	117	140

0.73 between memory and concentration, 0.54 between memory and intellect.

(18) *Dependence on race.* From studies conducted in several Missouri cities Pyle (61a) concludes that "in rote memory the negroes have a much better memory for concrete than for abstract words, but are greatly inferior to whites in both."

(19) *Miscellaneous observations.* (*a*) Reproduction in correct order is more difficult than mere reproduction; reproduction is more difficult than recognition.

(*b*) The first and the last terms of a series are more liable to be recalled than are the middle terms (Table 115).

(*c*) In word tests, certain terms are often found to have a special reproducibility, evidently by attracting special attention in some way. Thus, Binet and Henri found that the word *pupitre* (desk), though in the middle of a series, and hence unfavorably placed, was recalled in an unusually large number of cases.

(*d*) Errors of omission are more common than errors of insertion or errors of substitution—in word tests with school children, 4 times more frequent (Binet and Henri).

(*e*) Wissler calls attention to the perseverative tendency mentioned by Meumann and others: this is evinced by the introduction, in the recall of a given series, of impressions that had been used in an earlier series. Wissler found this type of error especially common in college seniors and mature *S*'s when trying the digit test. Meumann, it will be remembered, considered perseveration in the word test as an index of poor intelligence—when the *S*'s knew that no series was like a previous one.

REFERENCES

(1) Edwina E. Abbott, Memory consciousness in orthography. *Ps Mon*, 11: 1909 (No. 44), 127-158.

(2) A. R. Abelson, The measurement of mental ability of 'backward' children. *BrJPs*, 4: 1911, 268-314. (Same condensed as Tests for mental deficiency in childhood. *The Child*, 3: 1912, 1-17.)

(2a) E. J. Anderson, Standardization of the Heilbronner, rote memory and word-building tests. (Not yet published.)

(3) J. R. Angell, Methods for the determination of mental imagery, being pp. 61-107 in Rept. Com. Amer. Psychol. Assoc. on the Standardization of procedure in experimental tests. *PsMon*, 13: 1910. No. 53.

(4) I. M. Bentley, The memory image and its qualitative fidelity. *AmJPs*, 11: 1899, 1-48.

(5) J. A. Bergström, Experiments on physiological memory by means of the interference of associations. *AmJPs*, 5: 1893, 356-369. See also An experimental study of some of the conditions of mental activity, *ibid.*, 6: 1894, 267-273, and Relation of the interference and the practise effect of an association, *ibid.*, 6: 1894, 433-442.

(6) J. A. Bergström, Effect of changes in the time variables in memorizing, together with some discussion of the technique of memory experimentation. *AmJPs*, 18: 1907, 206-238.

(7) A. Bernstein and T. Bogdanoff, Experimente über das Verhalten der Merkfähigkeit bei Schulkindern. *Beiträge z. Psych. d. Aussage*, 2: 1905, 115-131.

(8) A. Binet, Attention et adaptation. *AnPs*, 6: 1899 (1900), 248-404.

(9) A. Binet, Introduction à la psychologie expérimentale. Paris, 1894. Especially ch. V.

(10) A. Binet and V. Henri, La mémoire des mots. *AnPs*, 1: 1894 (1895), 1-23.

(11) A. Binet and V. Henri, La psychologie individuelle. *AnPs*, 2: 1895 (1896), 411-465, especially 436-443.

(12) T. L. Bolton, The growth of memory in school children. *AmJPs*, 4: 1892, 362-380.

(13) B. Bourdon, Influence de l'âge sur la mémoire immédiate. *RPhF*, 38: 1894, 148-167.

(14) W. Brown, Some experimental results in the correlation of mental abilities. *BrJPs*, 3: 1910, 296-322. (Also embodied in his Use of the theory of correlation in psychology. Cambridge, Eng., 1910. Pp. 83.)

(15) W. H. Burnham, Memory historically and experimentally considered. *AmJPs*, 2: 1888-89, 39-90, 225-270, 431-464, 568-622.

(16) C. Burt, Experimental tests of general intelligence. *BrJPs*, 3: 1909, 94-177.

(17) C. Burt, Experimental tests of higher mental processes. *JEPd*, 1: 1911, 93-112.

(18) C. Burt and R. C. Moore, The mental differences between the sexes. *JEPd*, 1: 1912, 273-284, 355-388.

(19) Mary W. Calkins, A study of immediate and of delayed recall of the concrete and of the verbal. *PsR*, 5: 1898, 451-6.

(20) D. F. Carpenter, Mental age tests. *JEdPs*, 4: 1913, 538-544.

(21) W. G. Chambers, Individual differences in grammar grade children. *JEdPs*, 1: 1910, 61-75.

(22) J. Cohn, Experimentelle Untersuchungen über das Zusammenwirken des akustisch-motorischen u. des visuellen Gedächtnisses. *ZPs*, 15: 1897, 161-183.

(23) J. Cohn und J. Dieffenbacher, Untersuchungen über Geschlechts-, Alters- und Begabungs-Unterschiede bei Schülern. *Beihefte zur ZAngPs*, No. 2, 1911. Pp. 213.

(24) H. Ebbinghaus, (*a*) Ueber eine neue Methode zur Prüfung geistiger Fähigkeiten in ihrer Anwendung bei Schulkindern. *ZPs*, 13: 1897, 401-457. (*b*) For review of the literature, see *Grundzüge der Psychologie*, I, 2d ed., 1905, 633-707.

(25) E. Ebert and E. Meumann, Ueber einige Grundfragen der Psych. der Uebungsphänomene im Bereiche des Gedächtnisses. *ArGsPs*, 4: 1905, 1-232.

(26) P. Ephrussi, Experimentelle Beiträge zur Lehre vom Gedächtnis. Berlin, 1906. Pp. 191.

(27) J. Finzi, Zur Untersuchung der Auffassungsfähigkeit und Merkfähigkeit. *PsArb*, 3: 1900, 289-384.

(28) E. Fränkl, Ueber Vorstellungselemente und Aufmerksamkeit. Augsburg. 1905.

(29) H. Fuchs und A. Haggenmüller, Studien und Versuche über die Erlernung der Orthographie. *SmAbPdPs*, 2: 1898.

(30) F. Galton, Supplementary notes on 'prehension' in idiots. *Mind*, 12: 1887, 79-82.

(31) Eleanor Gamble, A study in memorizing various materials by the reconstruction method. *PsMon*, No. 43, Sept., 1909. Pp. 210.

(32) C. J. Hawkins, Experiments on memory types. *PsR*, 4: 1897, 289-294.

(33) V. Henmon, The relation between mode of presentation and retention. *PsR*, 19: 1912, 79-96.
(34) V. Henri, Éducation de la mémoire. *AnPs*, 8: 1901 (1902), 1-48.
(35) G. Heymans und H. Brugmans, Intelligenzprüfungen mit Studierenden. *ZAngPs*, 7: 1913, 317-331.
(36) H. Itschner, Lay's Rechtschrieb-Reform. *Jahrbuch des Vereins f. wissensch. Pädagogik*, 32: 1900.
(37) J. Jacobs, Experiments on 'prehension.' *Mind*, 12: 1887, 75-9.
(38) G. E. Johnson, Contribution to the psychology and pedagogy of feeble-minded children. *PdSe*, 3: 1895, 245-301, especially 268-273.
(39) W. F. Jones, An experimental-critical study of the problem of grading and promotion. *PsCl*, 5: 1911, 63-96, 99-120, especially 102-113.
(40) A. Jost, Die Associationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen. *ZPs*, 14: 1897, 436-472.
(41) F. Kemsies, Gedächtnisuntersuchungen an Schülern. *ZPdPs*, 2: 1900, 21-30, 84-95.
(42) E. A. Kirkpatrick, An experimental study of memory. *PsR*, 1: 1894, 602-9.
(43) Emma Kohnky, Preliminary study of the effect of dental treatment upon the physical and mental efficiency of school children. *JEdPs*, 4: 1913, 569-578.
(44) E. Kraepelin, Der psychologische Versuch in der Psychiatrie. *PsArb*, 1: 1896, 1-91, especially pp. 73 ff. (See also G. Aschaffenburg, *ibid.*, 209-299.)
(45) F. Krueger and C. Spearman, Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten. *ZPs*, 44: 1907, 50-114.
(46) F. Kuhlmann, A new memory apparatus. *PsR*, 19: 1912, 74-78.
(47) P. Lapie, Avancés et retardés. *AnPs*, 18: 1912, 233-270, especially 238-240.
(48) J. Larguier des Bancels, Sur les méthodes de mémorisation. *AnPs*, 8: 1901 (1902), 185-204.
(49) W. A. Lay, Experimentelle Didaktik. 3d ed., 1910, especially 297-305, 351-370.
(50) M. Lobsien, Experimentelle Untersuchungen über die Gedächtnissentwickelung bei Schulkindern. *ZPs*, 27: 1901, 34-76.
(51) E. Meumann, Intelligenzprüfungen an Kindern der Volksschule. *EPd*, 1: 1905, 35-101.
(52) E. Meumann, The psychology of learning. Eng. transl. New York, 1913. Pp. 393.
(53) W. McDougall, On a new method for the study of concurrent mental operations and of mental fatigue. *BrJPs*, 1: 1905, 435-445, especially 436 f.
(54) G. Müller and A. Pilzecker, Experimentelle Beiträge zur Lehre vom Gedächtniss. *ZPs, Ergänzungsband*, 1: 1900, 1-288.
(55) G. Müller and F. Schumann, Experimentelle Beiträge zur Untersuchung des Gedächtnisses. *ZPs*, 6: 1894, 81-190, 257-339.
(56) H. Münsterberg and J. Bigham, Memory. *PsR*, 1: 1904, 34-38.
(57) A. Netschajeff, Experimentelle Untersuchungen über die Gedächtnissentwickelung bei Schulkindern. *ZPs*, 24: 1900, 321-351.
(58) Naomi Norsworthy, The psychology of mentally deficient children (Columbia University thesis). New York, 1906. Pp. 111.
(59) M. Offner, Das Gedächtnis, Berlin. 1909. Pp. 238.
(60) A. Pohlmann, Experimentelle Beiträge zur Lehre vom Gedächtniss. Berlin, 1906. Pp. 191. (For full review, see *ZPs*, 44: 1907, 134-140.)
(61) W. H. Pyle, The examination of school children. New York, 1913. Pp. 70, especially 14-17.

(61a) W. H. Pyle, The mind of the negro child. *School and Society*, 1: 1915, 357-360.

(62) J. O. Quantz, Problems in the psychology of reading. *PsMon*, 2: 1897 (No. 5). Pp. 51.

(63) F. Reuther, Beiträge zur Gedächtnisforschung. *Psych. Studien*, 1: 1906, 4-101.

(64) C. Ritter, Ermüdungsmessungen. *ZPs*, 24: 1900, 401-444.

(65) M. C. Schuyten, Sur les méthodes de mensuration de la fatigue des écoliers. *ArPs(f)*, 4: 1904, 113-128.

(66) J. Segal, Ueber den Reproduktionstypus und das Reprodukzieren von Vorstellungen. *ArGsPs*, 12: 1908, 124-235.

(67) Stella E. Sharp, Individual psychology: a study in psychological method. *AmJPs*, 10: 1899, 329-391.

(68) B. R. Simpson, Correlations of mental abilities. *Columbia ConEd*, No. 53. New York, 1912. Pp. 122.

(69) W. G. Sleight, Memory training; is it general or specific? *JEPd*, 1: 1911, 51-54.

(70) F. Smedley, Report dept. child-study and pedagogic investigation (Chicago Public Schools). No. 3, 1900-1901, also in *RepComEd*, 1902, i., 1115-1138.

(71) W. G. Smith, The relation of attention to memory. *Mind*, n. s. 4: 1895, 47-73.

(72) W. G. Smith, The place of repetition in memory. *PsR*, 3: 1896, 21-31.

(73) W. G. Smith, A comparison of some mental and physical tests in their relation to epileptic and to normal subjects. *BrJPs*, 1: 1905, 240-260.

(74) Carrie R. Squire, Graded mental tests. *JEdPs*, 3: 1912, 363-380, 430-443, 493-506, especially 377-380.

(75) D. Starch, Transfer of training in arithmetical operations. *JEdPs*, 2: 1911, 306-310.

(76) A. v. Sybel, Ueber das Zusammenwirken verschiedener Sinnesgebiete bei Gedächtnisleistungen. *ZPs*, 53: 1909, 258-353. (For summary see *PsBu*, 7: 1910, 386-7.)

(77) E. B. Titchener, Experimental psychology, Vol. I., New York, 1901.

(78) J. J. van Biervliet, L'éducation de la mémoire à l'école. *RPhF*, 57: 1904, 569-586.

(79) R. Wessely, Zur Frage des Auswendiglernens. *Neue Jahrbücher f. Pädagogik*, 8: 1905, 297-309, 373-386.

(80) W. H. Winch, Immediate memory in school children. *BrJPs*, 1: 1904, 127-134, and 2: 1906, 52-57.

(81) W. H. Winch, (*a*) The transfer of improvement in memory in school children. *BrJPs*, 2: 1908, 284-293. (*b*) Same, continued, 3: 1910, 386-405.

(82) W. H. Winch, Mental fatigue in day-school children as measured by immediate memory. *JEdPs*, 3: 1912, 18-28, 74-82.

(83) C. Wissler, The correlation of mental and physical tests. *Ps Mon*, 3: 1901, No. 6. Pp. 62.

(84) S. Wyatt, The quantitative investigation of higher mental processes. *BrJPs*, 6: 1913, 109-133.

(85) C. S. Yoakum, An experimental study of fatigue. *PsMon*, 11: 1909, No. 46. Pp. 131.

TEST 39

Memory for ideas: 'Logical' memory.—This test differs from the preceding tests of memory in two respects: in the first place, connected, meaningful material is used instead of a series of disparate impressions; in the second place, the reproduction that is demanded is primarily a reproduction of ideas, not an exact, verbatim reproduction of the original presentation. In other words, this test, to use current phraseology, measures 'logical,' or 'substance' memory, instead of 'rote,' or 'mechanical' memory.

While, in principle, the attitude taken by *S* toward the test of memory for ideas is distinctly different from that taken toward the test of memory for discrete impressions, yet, in practise, it is not always possible to differentiate these attitudes in the tests as actually administered. Thus, Binet and Henri, and after them, Miss Sharp, conducted tests of "memory for sentences." In these tests, the sentences ranged from short to long, and from easy to difficult. A short, easy sentence, *e. g.*, a sentence of 11 words, is almost invariably interpreted by *S* as a straightforward test of verbal memory, and the reproduction is at bottom a recall in verbal (mainly auditory verbal) terms. On the other hand, a long, difficult sentence, *e. g.*, a sentence of 86 words, when heard or read but once, must be reproduced in substance, not verbatim, and the recall, for most *S*'s at least, is a recall by meaning, a reproduction of the 'gist' of the material presented.

It is evidently better to keep separate these two different forms of memory test, with their two correspondingly different attitudes. The material of the present tests is, accordingly, sufficiently lengthy to preclude verbatim recall. Memory for sentences of progressive length has been treated in Test 38.

The purposes of the test are similar to those of other memory tests, viz.: to determine individual differences in memory efficiency as related to sex, age, training, native ability, etc. As in those tests, too, the effect of different methods of presenting the material, or of different forms of material, may be studied, and immediate may be compared with deferred reproduction. The results of the test may also be correlated with the results of other tests, particularly with the tests of rote memory just described. Among others, the following examples are characteristic of these various uses of the logical memory test. In the Binet-Simon series of 1908 the Story of the Fire was introduced at the 8th and 9th years. Wissler used a logical memory

test in his series applied to Columbia freshmen, Terman in his comparative study of bright and dull boys, Winch in the investigation of fatigue and of transfer of training, Simpson in his comparison of competent and incompetent adults. The Marble Statue test, which the author derived from Shaw's study of memory in school children, has been applied by Pyle to several hundred children and adults for the purpose of establishing age and sex norms. Aall used an anecdote, much longer than those here prescribed, as the material for an interesting study of sex and individual difference, particularly as affected by immediate or deferred reproduction. Perhaps the most elaborate investigation of "memory for connected trains of thought" is, however, that of Henderson, who administered a series of tests to over 200 *S*'s, ranging from 10-year-old 5th-grade children to adult students in the university. Henderson's work forms the basis of the tests which are here prescribed, with some modifications suggested by the use of the test by the author for several years as a class exercise.

MATERIALS.—Watch. Three printed forms—The Marble Statue, Cicero, and The Dutch Homestead.

The first of these is taken from the appendix of Shaw's article, and was apparently used by him for subsidiary tests. The second and third are Nos. 2 and 4 of the five texts used by Henderson. If *E* wishes to extend the test by using more difficult material, he may employ Henderson's No. 5—a selection entitled "The Stages in the Development of Human Theory," from Comte's *Positive Philosophy*. If the Marble Statue proves too difficult or uninteresting for very young *S*'s, *E* may employ to advantage the text proposed by Binet and Simon (Wallin's arrangement) or the story of Mr. Lincoln and the Pig (from Clyde and Wallace, *Through the Year*, Book 2, Silver, Burdett & Co.).

THREE HOUSES BURNED

(From the 1908 Binet-Simon tests, revised by Wallin)

(51 words, 20 ideas)

New York, | September 5th. | A fire | last night | burned | three houses | in Water Street. | It took some time | to put it out. | The loss | was fifty thousand dollars, | and seventeen families | lost their homes. | In saving | a girl | who was asleep | in a bed, | a fireman | was burned | on the hands. |

How Mr. Lincoln Helped the Pig

(131 words, 42 ideas)

"One day | Mr. Lincoln | was out riding. | As he passed along the road, | he saw a pig | sinking | into a mud-hole. | Poor | piggy would climb | part way | up the slippery | bank, | then down he would fall again. |

'I suppose I should get down | and help | that pig,' | thought Mr. Lincoln. 'But I have on my new suit, | and it will be quite spoiled if I do so. | I think I'll let him get out | the best way he can.' |

He rode on. | When nearly | two | miles away, | he turned | and came back. | Not minding the new | clothes, | he stooped, | and taking piggy in his arms, | he dragged him | out | of the mud. |

The new | suit | was quite | spoiled, | but Mr. Lincoln | said | he had taken a pain | out of his mind." |

Method.—Provide *S* with paper and pencil. Explain the nature of the test, as follows: "I am going to read you something to see how well you can remember it afterward. You must pay careful attention, as I shall read it but once. As soon as I have finished, take your pencil and write as much of the story as you can remember. If you can remember it in just the words you heard, use those words, but if you can't do that, tell in your own words, as well as you can, what it was that I read to you."

Read the passage, including the title, with most careful enunciation, and with proper attention to expression. The rate of reading should be somewhat slower than in ordinary reading—say a full minute for the Cicero test. Allow *S* ample time for writing, then ask him to underline each word in his reproduction that he feels sure is exactly the same as the original passage.[1]

[1]This test lends itself easily to the group method. The usual precautions should be taken to avoid disturbance and communication. *E* may save himself much labor by asking each *S* to count the total number of words he has written, then the total number of words he has underlined. With mature *S*'s, *E* may also reread very slowly the original text, and let each *S* check up the total number of ideas correctly reproduced, *i. e.*, represented, whether verbatim or by equivalent phrases, in his reproduction. The division of each text into its constituent 'ideas' is indicated below.

Variations of Method.—(1) Supply *S* with the printed text. Inform him that he is to have 2 min. to read the passage. Assure him that this time is ample to read it over carefully more than once. Direct him to read the passage straight through twice, and then use any time that remains in studying it as he wishes.

(2) Defer the reproduction to any desired time after the reading, *e. g.*, 10 min., 24 hours, 1 week, 4 weeks. Or require an immediate reproduction, followed later, at one or more of the intervals just suggested, by a second or by a third reproduction.[2] Conduct these deferred trials in the same manner, as far as directions to underline, etc., are concerned, as in the first trial.

Treatment of Data.—The simplest plan for scoring the data of this test is that used by Terman and by Wissler, who merely graded the papers on a scale of 5 (or of 10) for a perfect reproduction—perfect in the sense of a reproduction of all the ideas of the original text, whether in terms identical with, or merely equivalent to, the original.

For ordinary purposes, the author has found it serviceable to score the papers for the following points: (1) number of words written, (2) number of words underlined, (3) percentage of underlined words that are correctly underlined, (4) number of ideas ('details' in Henderson's terminology) that have been reproduced, whether exactly or in equivalent phrases. To these may be added, if desired, (5) number of ideas wrongly inserted. If but a single score is to be made, the fourth is obviously the one to be used, since the task assigned to *S* is to give as many as possible of the ideas of the text.[3]

The second and subsequent reproductions are scored in the same manner as the first. Retention is then measured, follow-

[2]It is better, on the whole, to give no intimation of the intent to demand a second reproduction. Some *S*'s may compare notes after the first reproduction, but if the subsequent trial is announced beforehand, coupled, as it ought to be, with the request not to think of the test in the interim, the request is more apt to work as a counter-suggestion, so that many *S*'s will test their recall of the passage, and otherwise furbish up the memories during the interval.

[3]Consult Aall, Sharp or Henderson for more elaborate methods of treating data, particularly for devices for qualitative analysis.

ing Shaw and Henderson, by computing the percentage of loss between these and the first reproduction. Occasional cases of improvement in the later reproductions are rated as a negative loss.

To ascertain the 'idea-score,' *S*'s reproduction must be compared, step by step, with the standard divisions of the original text into ideas.[4]

The Marble Statue

(166 words, 67 ideas)

A young | man | worked | years | to carve | a white | marble | statue | of a beautiful | girl. | She grew prettier | day by day. | He began to love the statue | so well that | one day | he said to it: | "I would give | everything | in the world | if you would be alive | and be my wife." | Just then | the clock struck | twelve, | and the cold | stone began to grow warm, | the cheeks red, | the hair brown, | the lips to move. | She stepped down, | and he had his wish. | They lived happily | together | for years, | and three | beautiful | children were born. | One day | he was very tired, | and grew | so angry, | without cause, | that he struck her. | She wept, | kissed | each child | and her husband, | stepped back | upon the pedestal, | and slowly | grew cold, | pale | and stiff, | closed her eyes, | and when the clock | struck | midnight, | she was a statue | of pure | white | marble | as she had been | years before, | and could not hear | the sobs | of her husband | and children.

Cicero

(125 words, 64 ideas)

"Cicero, | the greatest | of the Roman | orators, | was born | at Arpinum, | an obscure | country | town. | His family | was of the middle class | only, | and without wealth, | yet he rose | rapidly | through the ranks | of Roman | official service | until at the age | of forty-six | he became | consul. | In oratory | he

[4]The scoring for ideas for these three passages is taken, with a few minor changes, from Shaw and from Henderson. For a division of the second and third texts into topics and sub-topics as well as into ideas, the reader may consult Henderson (6, pp. 29-30).

is | by universal consent | placed side by side | with Demosthenes, | or at least | close after him. | He surpassed | the great | Attic | orator | in brilliancy | and variety, | but lacked | his moral | earnestness | and consequent | impressiveness. | He could be | humorous, | sarcastic, | pathetic, | ironical, | satirical, | and when he was malignant | his mouth was | most | foul | and his bite | most | venomous. | His delivery | was impassioned | and fiery, | his voice | strong, | full, | and sweet, | his figure | tall, | graceful, | and impressive."

THE DUTCH HOMESTEAD

(180 words, 94 ideas)

"It was | one | of those spacious | farm- | houses, | with high- | ridged, | but lowly | sloping | roofs, | built | in the style | handed down from | the first | Dutch | settlers, | the low | projecting | eaves | forming a piazza | along the front | capable | of being closed up | in bad weather. | Under this | were hung | flails, | harness, | various | utensils | of husbandry, | and nets | for fishing | in the neighboring | river. | Benches | were built | along the side | for summer use; | and a great | spinning wheel | at one end, | and a churn | at the other, | showed | the various | uses | to which this | important | porch | might be devoted. | From this piazza | one might enter | the hall, | which formed | the center | of the mansion | and the usual | place of residence. | Here | rows | of resplendent | pewter | ranged | on a long | dresser | dazzled | his eyes. | In one corner | stood a huge | bag | of wool, | ready | to be spun; | in another | a quantity | of linsey-woolsey, | just | from the loom; | ears | of Indian | corn | and strings | of dried | apples | and peaches | hung | in gay | festoons | along the walls, | mingled | with the gaud | of red | peppers."

RESULTS.—(1) *Norms.* Performance with the Marble Statue selection has been investigated by Pyle (group method, written reproduction) with the results shown in Table 116. The author's results for college students with the Dutch Homestead selection are shown in Table 117, while some idea of the dis-

tribution of performance for the Cicero selection with college students (mostly Sophomores) is given by Table 118. In all of these tables, and particularly in those referring to college students, evidence is given to show the unexpectedly large individual variation in memory for ideas that prevails even within a group of *S*'s of apparently similar attainments.

TABLE 116

Marble Statue Test, Scores by Age and Sex (Pyle)

SEX		8	9	10	11	12	13	14	15	16	17	18	ADULT
Male	Cases	102	148	142	149	156	163	129	89	60	45	32	65
	Aver.	24.3	28.7	30.0	32.9	35.1	36.8	36.1	36.5	34.4	34.6	36.9	38.3
	A. D.	6.7	9.1	6.7	5.6	7.4	6.3	7.0	6.7	5.6	8.7	6.0	7.0
Female	Cases	89	158	138	156	191	164	146	99	94	81	48	86
	Aver.	28.5	31.0	33.5	36.4	38.1	38.5	39.0	39.1	37.3	36.6	37.8	40.1
	A. D.	11.3	9.4	6.8	7.7	7.2	7.1	7.5	6.3	5.1	6.9	4.4	5.9

TABLE 117

Dutch Homestead Test. Words Written and Underlined (Whipple)

	FIRST TRIAL, NO INTERVAL		SECOND TRIAL, 24 HOURS LATER	
	Total Words	Words Underlined	Total Words	Words Underlined
Average, 9 men	80.4	48.0	83.0	38.4
Average, 22 women	95.5	38.8	99.6	34.1
Maximal records	127.0	102.0	138.0	66.0
Minimal records	45.0	4.0	52.0	9.0

TABLE 118

Cicero Test. Distribution of 36 College Students (Whipple)

	BEST SIX	2D SIX	3D SIX	4TH SIX	5TH SIX	WORST SIX
Words written	115-96	95-88	84-75	74-66	65-60	56-25
Correct ideas	46-37	36-34	32-28	28-25	25-21	20-7

(2) *Dependence on age.* Binet and Simon put the reproduction of 2 items from the Story of the Fire as standard for 8 years, 6 items as standard for 9 years. The results reported by Vos, who read to boys and girls 9 to 14 years old a story containing 40 'ideas' and called for reproductions 3 days later, show rather unusual variations from year to year, despite the fact that some 800 S's are represented: report is declared to be good at 9, best at 10, thence deteriorating decidedly to 13, but improving at 14 years. The elaborate studies of Shaw and of Henderson are also somewhat difficult to interpret. It appears evident, however, that a distinction must be made between efficiency in the first reproduction and efficiency in subsequent reproductions. If the first be termed learning capacity, and the second retentive capacity, and if the latter be measured in terms of the proportion of the first reproduction that is retained in the second (or later) reproduction, then adults may be shown to surpass children in learning capacity, but not in retentive capacity.

Thus, in Shaw's rather difficult 324-word story, the learning capacity of boys increased, from the 3d to the 9th grade, from 17 to 42 per cent., that of girls from 18 to 43 per cent. High-school boys averaged only 40 per cent., high-school girls about 47 per cent. Shaw's university students did no better, while Henderson's summer session students were inferior to his 15 and 16-year-old school children. In short, then, logical, like rote memory appears, when measured by the first reproduction, to be at its best near puberty. This conclusion agrees entirely with the averages secured by Pyle for children 8 to 18 years old and for adults (Table 116). His boys reached their maximal ability at 13, his girls at 12 years.

TABLE 119

Average Percentage of Loss in Third Reproduction (After 4 Weeks) (Henderson)

AGE	ADULTS	16	15	14	13	12	11	10
Percentage of loss	14	8	13	15	14	12	10	10

Turning to the later reproductions, Shaw and Henderson (Table 119) agree that younger *S*'s have as good retentive capacity as do adults.

(3) *Sex differences* in this test, as in the rote-memory test, are in favor of girls. The difference is indicated clearly in the author's data for college students (Table 117), and similar differences are reported by Shaw, who found the growth of memory for ideas to be faster in girls than in boys, and the average performance of girls to be some 4 per cent. better than that of boys. Wissler's records for Columbia freshmen show an average of 44.5 per cent., P.E. 11.1, for men, and 48.2 per cent., P.E. 13.2, for women. Pyle's averages reveal the superiority of girls at every age from 8 to maturity. Schramm's comparison of 16 men and 16 women (students at Freiburg University) shows a slight superiority for the women, though the differences do not exceed their probable error. The only exceptions to this trend in favor of superiority of females seem to appear in the work of Vos and of Aall. The latter states that the reproductions of women are usually fuller, but those of men are more compact, 'meatier,' and betray greater plastic power, greater originality in formulation.

(4) *Dependence on time-interval.* The insertion of a time-interval between presentation and reproduction has much less effect upon memory for ideas than upon memory for discrete impressions. Table 117 shows that, if a second reproduction is called for one day after the first, the average *S* actually writes more words. The words in the later reproduction are, however, less exact copies of the original text, and there is a tendency to insert extraneous material, so that fewer words are underlined, and there is a slight net reduction in the number of ideas reproduced. In the author's tests, this reduction was but 3 per cent. at the end of one week. Table 119 shows that an interval of 4 weeks produces a loss of but 8 to 15 per cent.

Similar conclusions are reached by Aall, who compared the reproductions of a lengthy story directly after hearing it and 48 hours later. The latter reports were on the whole poorer—more omissions and more 'falsification.' They are shown to 'lean'

strongly on the first reports, often to become more verbose and less precise, and sometimes matters that were correctly left rather vague in the first report become erroneously 'logicized' into explicit and particularized statements, which are actually possible, but incorrect as reports. On the other hand, as Aall points out, there exists a sort of 'after-memory' such that certain details which are forgotten or at least unmentioned in the immediate reproduction come to light correctly in the delayed one. The influence of time-interval also appears to operate differently on different forms of material; for instance, names of places are lost sooner than memories of objects.

It is a matter of special interest to note that the relative standing of *S*'s remains practically the same in tests conducted with immediate, and with deferred recall. Similarly, those who memorize a passage of a given length in quick time are not found to be at a disadvantage in subsequent recall (see Ogden and Pyle, 12), though the individual differences are usually found to be less in subsequent recall than in original speed of learning. It follows that, so far as this test goes at least, the popular notion "easy come, easy go" is not borne out by experimental evidence. Henderson found that this correlation between learning capacity and retentive capacity was brought out better in scoring for ideas than in scoring for words.

(5) *Dependence on method of presentation.* When a single hearing is compared with reading done by *S* (3 min.), the former is found to be nearly as good as the latter for immediate reproduction, but the latter to be much more effective than the former for deferred reproduction.

(6) *Dependence on practise.* Baade, who scored the reports made by 196 girls, aged 12-13 years, upon what was said by an instructor in the course of a series of demonstrations in physics, found that under those conditions there was no demonstrable improvement in their work, either from the succession of the three sets of demonstrations or from the repetition of the demonstrations. Nevertheless, the work of other investigators gives little doubt that practise will improve memory for ideas, as it will improve nearly every form of psychophysical activity. Special training thus accounts, in all probability, for the high

scores (52 as over against 40 to 47 per cent.) reached by the pupils of Miss Aiken's school[5] in comparison with the work of Worcester high-school children.

Winch (21) has investigated the possibility of transfer of practise, and concludes that "improvement through practise in rote memory for things with and without meaning is followed by improvement in substance for stories." He argues that this transfer may take place despite the circumstance that correlations between rote and substance memory are sometimes, as in his own work, of a low or even doubtful character. Winch (23) has also investigated the transfer of practise in substance memory to efficiency in productive imagination. He concludes that "children practised in substance memory for stories become thereby more proficient in the invention of stories. The improvement is not due to the insertion of parts of the content of the memorized stories within the invented stories, but to some community of function less atomistic." An exception appeared, however, in portions of his experimental work, which leads him to add that "children practised in substance memory up to the fatigue point, which is taken here to mean the point at which consecutive exercises cease to produce improvement, are thereby prejudicially affected so far as their power to invent stories is concerned." These 'fatigue-effects,' he says, "appear to be temporary, whilst practise effect (improvement through practise) appears to have considerable duration."

(7) *Dependence on fatigue.* Although it is generally conceded that one of the commoner symptoms of mental fatigue is slowness or uncertainty of recall of ideas, no one but Winch (22) appears to have used the logical memory test in this connection. Winch's test was limited to a group of boys who were studying at an evening school. They were given 10 minutes to memorize the substance of passages of some 150 words and tested by the method of equivalent groups at 9 and at about 9.30 P. M., with the result that some 28 per cent. of difference was revealed. Since tests in day schools had shown prac-

[5]For an account of the special training given to Miss Aiken's pupils, see Test 25 and references thereto.

tically no difference between children working at noon and at 4.30 P. M., Winch argues that children who take evening school work after a day's work at various occupations exhibit a very rapid and pronounced susceptibility to mental fatigue.

(8) *Dependence on length of text.* The number of words reproduced after one hearing increases, though not in direct proportion, with the length of the passage heard (Binet and Henri).

(9) *Dependence on portion of text.* If the original passage be divided into 3ds or 4ths (or even, if long, into 8ths), it will be found that, on the average, the reproduction of any one of these portions is inferior to the one that precedes it and superior to the one that follows it. Thus, Shaw's story, on division into 4 parts, was found to be reproduced in the amounts 52, 34, 31, and 28 per cent., respectively.

What may be regarded as a test of logical memory was made by Dell with 30 boys who listened to an hour-and-a-half lecture upon material in *Punch*, which was illustrated by 80 lantern slides. The boys were asked 8 days later to indicate which slides they recalled. The first 15 slides were, on the average, recalled by 11.2 boys, slides 47-59 by 3.8, slides 60-74 by 3.5, and slides 75-80 by 4.8 boys. These figures certainly seem to demonstrate a decidedly better recall of slides in the first portion of the lecture, with a questionable slight rise at the end of the lecture. Dell, however, believes that there were at work other causes than simply the dependency of memory on different positions in the lecture.

(10) *Reliability.* The work of Simpson, Winch and others shows that the logical memory test has an acceptable degree of reliability. Winch obtained coefficients of 0.65 and 0.68 between single trials; Simpson, coefficients of 0.78, 0.83 and 0.90 (for different groups) between scores in his first two and his last two trials. It follows that the amalgamated results from two, or at most three trials of this test afford quite reliable indications of ability in the capacities tested.

(11) *Dependence on intelligence.* The relation between logical memory and intelligence has been studied mainly by rating intelligence on the basis of scholarship. The resulting correla-

tions are somewhat unexpectedly low. Thus, in 86 cases, Wissler found a correlation of only 0.19 with class standing, of 0.11 with standing in mathematics, and of 0.22 with standing in Latin. Henderson found but a slight correlation with class standing in the lower grades, but a closer correlation in the higher grades. He is of the opinion that, at least in the lower grades, the school marks put a premium upon industry and good conduct, rather than upon native ability, and thus obscure the existing correlation. Pyle estimates the relation with class standing in college at about 0.30, and says: "If a slow learner has the habit of going over a lesson or task several times, and a fast learner the habit of giving a lesson but one hasty reading, other things being equal, the slow learner will have the better scholarship" (12, p. 319). The very best students have both good memory and good habits of study. If performance in tests of reasoning be taken as a measure of intelligence, then Peterson's work confirms the general statement given above, for of 30 students classed as good in reasoning, 20 ranked good, 5 medium and 5 poor in memory, while of the 17 classed as poor in reasoning 3 ranked good, 6 medium and 8 poor in memory. Again, in Simpson's investigation the correlation between logical memory and estimated intelligence of his 'good' group was but 0.35, after correction for attenuation. However, his good and his poor group were fairly well separated by the test, since none of the poor group reached the median of the good group and only 15 per cent. of the poor group excelled the lowest 12 per cent. of the good group.

(12) *Mental defectives.* Wallin tested epileptics by means of Binet's Story of the Fire, though without warning them when reading that a reproduction would be called for. Table

TABLE 120

Story of the Fire. Scores for Epileptics by Mental Age (After Wallin)

BINET-SIMON AGE	VII	VIII	IX	X	XI	XII	XIII
Cases	13	42	27	70	28	11	17
Aver. Ideas	2.1	3.7	4.8	5.3	6.5	6.8	7.4

120 shows the average number of ideas reproduced by his patients as classified for mental age by the Binet-Simon tests. It is evident that there exists a general progress in efficiency with mental age, but that the average performance of these mental defectives is not as good as would be expected of normal children in the ages from 9 to 13.

(13) *Miscellaneous correlations.* Peterson, by the method of unlike signs, using the pooled results of several tests of each 'function,' concluded that memory was correlated with reasoning by 0.40, with abstract thought by 0.64; with generalizing ability by 0.40 and with accuracy by 0.31. Simpson publishes as estimated true correlations (holding for people in general) with memory for passages the following: completion test 0.71, hard opposites 0.70, memory for words 0.80, easy opposites 0.50, a-test 0.46, adding 0.42. The high correlation with memory for words permits us, he concludes, to class substance memory with it as virtually the same capacity. Winch found correlations between substance memory and productive imagination (inventing stories) which appear to be higher in the more proficient classes (the r's secured in various classes were 0.28, 0.43, 0.48, 0.62, 0.75). Heymans and Brugmans found a correlation of only 0.08 between reproduction of a somewhat elaborate story at various time-intervals and the learning of nonsense syllables, but a correlation of 0.55 between the memory test and a test involving memory for details of a picture. They also report a correlation of 0.56 between the completion test and written reproduction of a difficult passage from Höffding's Ethics. Wissler found a correlation of 0.21 between logical memory and length of head, but no correlation between logical memory and rote memory, speed of naming colors, reaction time, or breadth of head.

(14) *Qualitative aspects.*[6] Inspection of the work of children and introspective examination by adults of their own mental

[6]Consult especially Balaban, Michotte and Ransy, and Michotte and Portych for further study of the qualitative aspects of logical memory, particularly of the difference between mechanical and logical memory under simple test conditions, like the method of right associates. The work of Aall has also numerous suggestive features in addition to those here mentioned.

processes reveal a number of interesting principles. In the first place, there is a process of selection: words or ideas that are logically or psychologically important are best retained. Or, as Henderson expresses it, there is, especially during a long time-interval, a process of condensation and generalization. The main ideas, the important topics, the brunt of the passage may remain fairly constant, but the minor details tend to be forgotten, and the original phrasing to become less and less clear.

When, then, the reproduction is demanded, most *S*'s first recall these main ideas or larger topics, and then develop the details, as best they may, from them. There is a strong tendency, in this filling out of the details, toward what Binet and Henri speak of as "verbal assimilation," *i. e.*, a tendency to express the ideas in one's own terms, rather than in those employed in the original passage. Thus, adults often use synonyms or other forms of substitution, while children replace the words of the original by words from their ordinary vocabulary (*e. g., played* for *amused themselves, fire* for *conflagration*), and at the same time tend to simplify the syntax. In general, Binet and Henri found that the number of times that synonyms are used in the recall is, in short passages greater, and in long passages less than the number of ideas completely omitted.

Finally, the substitution of terms for those of the original tends, especially in younger *S*'s and with longer time-intervals, to become inexact; in other words, the sense of the original becomes more or less distorted. Thus, for instance, Binet and Henri discovered that, in all sentences containing more than 20 words, more than half of their *S*'s had made some change in the meaning of the original. Of these alterations of sense, the most conspicuous are: (1) change of proper names or of numbers, (2) replacement of an object by an analogous object that might fit the sentence equally well, (3) insertion of details not inconsistent with the original, but still not in the original, and (4) alterations apparently due to emotional reaction, especially to exaggeration, *e. g., a frightful snake* for *a snake*.

Aall thinks that two sorts of *S*'s can be distinguished—the 'reporters,' who make every effort to get the reproduction ex-

act, and the 'describers,' who introduce various modifications and embellishments for the sake of literary or rhetorical effect. He found the typical error in recounting a story to be omission rather than falsification. Mention has been made already of the changes found by Aall in deferred reproductions.

(15) *Miscellaneous points.* In the case of college students, from 50 to 90 per cent. of the words underlined are actually correct. A certain type of *S* may be recognized, who is extremely cautious about underlining words, but who usually has these few nearly all correct.

Binet and Henri estimate that memory for connected sentences is approximately 25 times as good as memory for discrete terms.

REFERENCES

(1) A. Aall, Zur Psychologie der Wiedererzählung. *ZAngPs*, 7: 1912-13, 185-210.

(2) W. Baade, Aussage über physikalische Demonstrationen. I. Abh. Die Methodik der Versuche über die Inhalte der Textaussagen. *ZAngPs*, 4: 1911, 189-311.

(3) A. Balaban, Ueber den Unterschied des logischen und des mechanischen Gedächtnisses. *ZPs*, 56: 1910, 356-377.

(4) A. Binet et V. Henri, La mémoire des phrases. *AnPs*, 1: 1894 (1895), 24-59.

(5) J. A. Dell, Some observations on the learning of sensible material. *JEdPs*, 3: 1912, 401-406.

(6) E. N. Henderson, A study of memory for connected trains of thought. New York. Pp. 94. (Columbia Univ. Thesis.)

(7) G. Heymans und H. Brugmans, Intelligenzprüfungen mit Studierenden. *ZAngPs*, 7: 1913, 317-331.

(8) A. Michotte et C. Ransy, Contribution a l'étude de la mémoire logique. *Extrait des Annales de l'institut supérieur de philosophie.* Louvain, 1912. Pp. 95.

(9) A. Michotte et Th. Portych, Deuxième étude sur la mémoire logique. *Ibid.*, 2: 1913, 533-657.

(10) R. M. Ogden, Untersuchungen über den Einfluss der Geschwindigkeit des lauten Lesens auf das Erlernen und Behalten von sinnlosen und sinnvollen Stoffen. *ArGesPs*, 2: 1903-4, 93-189.

(11) H. A. Peterson, Correlation of certain mental traits in normal school students. *PsR*, 15: 1908, 323-338.

(12) W. H. Pyle, Retention as related to repetition. *JEdPs*, 2: 1911, 311-321.

(13) W. H. Pyle, The examination of school children. New York, 1913. Pp. 70, especially 8-14.

(14) F. Schramm, Zur Aussagetreue der Geschlechter. *ZAngPs*, 5: 1911, 355-357.

(15) Stella E. Sharp, Individual psychology: a study in psychological method. *AmJPs*, 10: 1899, 329-391.

(16) J. C. Shaw, A test of memory in school children. *PdSe*, 4: 1896, 61-78.

(17) B. R. Simpson, Correlations of mental abilities. *Columbia ConEd*, No. 53. New York, 1912. Pp. 122.
(18) L. W. Terman, Genius and stupidity. *PdSe*, 13 : 1906, 307-373.
(19) H. B. L. Vos, Beiträge zur Psychologie der Aussage bei Schulkindern. Analyse der Aussage über eine gehörte Erzählung. Amsterdam, 1909. (Eigenbericht, *ZAngPs*, 4 : 1911, 375-378.)
(20) J. E. W. Wallin, Experimental studies of mental defectives. *EdPsMon*, No. 7, 1912. Pp. 155.
(21) W. H. Winch, The transfer of improvement in memory in school children, II. *BrJPs*, 3 : 1910, 386-405.
(22) W. H. Winch, Some measurements of mental fatigue in adolescent pupils in evening schools. *JEdPs*, 1 : 1910, 13-23, 83-100.
(23) W. H. Winch, Some relations between substance memory and productive imagination in school children. *BrJPs*, 4 : 1911, 95-125.
(24) C. Wissler, The correlation of mental and physical tests. *Ps Mon*, 3 : 1901. Pp. 62.

CHAPTER X

Tests of Suggestibility

The term 'suggestion' has found different usages in psychology; four at least may be readily distinguished. (1) Suggestion is equivalent to association, *e. g.*, the idea 'horse' suggests the idea 'Black Beauty.' (2) Suggestion is the conveyance of an idea by hint, intimation, or insinuation, *e. g.*, the orator suggests an idea by an appropriate gesture. (3) Suggestion is a method of creating and controlling hypnosis. (4) Suggestion is a process of creating belief or affecting judgment, usually an erroneous belief or false judgment, in the normal consciousness. Here emphasis is placed upon uncritical acceptance of a notion usually with the implication that the suggested individual is unaware that his ideas have been thus affected. From his point of view, suggestion is, then, to follow Stern's definition (6), "the imitative assumption of a mental attitude under the illusion of assuming it spontaneously."

The tests which follow all purport to measure susceptibility to suggestion in this last-named sense. In them, the experimenter seeks, by suitable arrangement of the test-material or of the instructions, to induce the subject to judge otherwise than he naturally would—to induce him, for example, to judge equal lines or equal weights to be unequal, or to perceive warmth when there is no warmth, etc. If the attempt is successful, the subject is said to have 'yielded,' or to have 'accepted' the suggestion; if unsuccessful, he is said to have 'resisted' the suggestion. The degree of his suggestibility is indicated by the quickness or frequency of his 'yields.'

Just as efficiency in observation, attention, memory, and the like has been shown to be specific, not general, in character, so is it probable that suggestibility is specific, not general, in character. For this reason, suggestibility must be tested by more than one method.

Many of the tests in other portions of this book, *e. g.*, Nos. 17, 23, and especially 32, afford opportunity for noting the suggestibility of subjects. The serial graded tests of Binet and Simon also contain directions for testing the suggestibility of young or of feeble-minded children.

The tests which follow deal with suggestibility aroused by the volume-weight illusion, the length of lines, judgments of weights under special conditions, and illusory warmth. Other experimental methods of inducing suggestibility, which have not as yet been arranged for test work, may be briefly cited.

H. J. Pearce (3) had *S*'s sit in a chair with a circle of 3.5 ft. radius drawn about it. *S* fixated a small bit of paper directly in front of him. A test square was exposed briefly at a point somewhere to the right of the fixation-point, and *S* located its position afterward by moving his eyes to the right. Suggestion was introduced by displaying at times a third bit of paper near or farther than the test square. There was at first a tendency to resist this suggestion, but eventually there was developed a tendency to locate the test square in a direction corresponding to the location of the suggestive paper. Auditory and tactual stimuli were also tried.

J. C. Bell (1) displayed triangles of different shapes and heights, also vertical distances between points or between a point and a line. The *S*'s reproduced the distances and were given verbal suggestions or visual suggestions to "make high" or "make low," etc. In general, the suggestions did affect the test with triangles, but there were decided individual differences, and in many cases the constant errors were greater than the errors induced by suggestion.

The work of E. K. Strong (7) was similar in character, save that his *S*'s exerted maximal strength of grip while exposed to such suggestions as "Now you can make it stronger than usual," etc., but with the proviso not voluntarily to interfere with the suggestion. The results showed that grips following both suggestions of 'weak' and 'strong' were stronger than those intended to produce 'neutral,' while there was no difference between two first kinds.

W. D. Scott (5) produced suggestion with some success by inducing *S*'s to think that the flight of colors following an exposure to white light corresponded in order to the arrangement of colors in the spectrum.

Inez Powelson and M. F. Washburn (4) showed colors with comments upon them, such as 'delicate,' 'crude,' etc., and influenced in this way the affective reactions of 19 *S*'s, but failed with 16 others.

Giroud (2) showed 34 children, aged 7 to 12 years, a series of 10 colors, with the instructions to name each color and then, when the color was withdrawn, to write the name. At the 3d, 7th and 10th terms, a wrong color-name was uttered by the experimenter to try to induce *S* to write it. The average number of 'yields' was reduced gradually from 2.8 at 7 years to 1.7 at 12 years.

REFERENCES

(1) J. C. Bell, The effect of suggestion upon the reproduction of triangles and of point distances. *AmJPs*, 19: 1908, 504-598.

(2) A. Giroud, La suggestibilité chez des enfants d'école de sept à douze ans. *AnPs*, 18: 1912, 362-388.

(3) H. J. Pearce, Normal motor suggestibility. *PsRev.*, 9: 1902, 348-355.

(4) Inez Powelson and M. F. Washburn, The effect of verbal suggestion on judgment of the affective value of colors. *AmJPs*, 24: 1913, 267-9.

(5) W. D. Scott, Personal differences in suggestibility. *PsRev*, 17: 1910, 147-154.

(6) W. Stern, Abstracts of lectures on the psychology of testimony. *AmJPs*, 21: 1910, 270-275.

(7) E. K. Strong, The effect of various types of suggestion upon muscular activity. *PsRev*, 17: 1910, 279-293.

TEST 40

Suggestion by the size-weight illusion.—Big things are ordinarily heavier than small things of the same kind. When we lift two weights of apparently the same material, but of different sizes, we more or less unconsciously put forth more energy or expect to meet with more resistance in lifting the larger. If, as in the case of the so-called 'suggestion-blocks,' the weights are really the same, we almost inevitably judge the larger weight to be the lighter; in other words, the visual appearance

of the weight has given us a suggestion—or, as it turns out, rather, a disappointed suggestion—of weight.[1]

This error of judgment is undoubtedly due to an association built up by long experience in handling and lifting various articles and objects.[2] One might, therefore, suppose that younger children, or less intelligent children, who would, presumably, have had less of this discriminative association of size and weight, would be less affected by the suggestion. For this reason, the size-weight test has been applied by several investigators to determine or to measure, at least relatively, the degree of suggestibility exhibited by school children under various conditions. But it is to be noted that the having of the illusion is normal, so that this test is not on the same order as those that follow it, and it has probably no particular value as a measure of suggestibility in older children and adults; its primary value lies in its use with young or mentally defective children.

APPARATUS.—Low table. Soft black cloth. Set of 'suggestion-blocks,' patterned after Gilbert, but modified by extending the comparison series in both directions.

This set consists of two standard blocks and 20 comparison blocks. Both standards weigh 55 grams; both are 28 mm. thick, but the larger is 82 and the smaller 22 mm. in diameter. The 20 comparison blocks are all 28 mm. thick and 35 mm. in diameter, but their weights range from 5 to 100 g. by 5 g. increments.[3] All are painted dead black.

If it is desired merely to make a quick determination of the presence or absence of the illusion, simpler material may be employed, preferably the 'Demoor blocks' as used at the Vineland (N. J.) Training School. These blocks are of poplar wood, and both weigh 1.5 lbs.; the one is 1.75x3x4 in., the other 1.75x4x12 in. They are set before *S* with the simple instruction: "Tell me which block seems the heavier."

METHOD.—Arrange the table at such a height that *S*'s fore-

[1]As Scripture remarks, the poor fellow who has been laughed at for centuries for saying that a pound of lead is heavier than a pound of feathers is perfectly right, so long as he speaks psychologically, and looks at the pillow and the bit of lead pipe. A concrete demonstration of the truth of this statement is afforded by several experiments reported by Wolfe.

[2]Some writers, however, *e. g.*, Flournoy, attribute the illusion to an inborn nervous connection. For a discussion of the psychological factors concerned in this experiment, particularly in its relation to the 'innervation-sense,' consult Flournoy, Müller and Schumann, Seashore, Bolton, Loomis, and van Biervliet.

[3]Gilbert's comparison blocks were but 14 in number, with a range from 15 to 80 g. This range proved inadequate for younger *S*'s.

arm will be parallel with the floor when lifting a weight. Spread over the table the black cloth, which should be large enough to cover at least the portion of the table occupied by the weights, and thick enough to deaden the sounds incident to their replacement.

Arrange the twenty comparison blocks on the cloth, in the order of their weight from left to right, and in such a manner that any one of them may be reached by *S* without materially changing the angle of his arm. Place before *S* the larger standard block, and say: "Here is a block. I want you to find a block in this series of 20 blocks that seems to you just as heavy as this one. Lift it by picking it up edgewise with your thumb and finger, like this. [Illustrate.] Then try the first of these weights [at the left]. If that doesn't suit, try the next, then the third, and so on, till you find a block that seems equal to this one. Each time you must lift this block first, then the one you are trying in the series. Keep your eyes constantly directed at the weight you are lifting." When *S* has selected an equivalent weight, the same procedure is followed with the second, or smaller, standard block.

The work of investigators in the psychological laboratory, particularly Martin and Müller, and Müller and Schumann, has shown that our estimate of the absolute or relative weight of a body is conditioned by an unsuspectedly large number of factors, so that, while it may be true, as Fourche (8) asserts, that voluntary modifications in the speed of grasping and lifting the weights do not modify the size-weight illusion, it seems desirable that the conditions under which *S* lifts the blocks should be kept as uniform as possible.

S should pick up each block in the same manner, lift it at the same tempo and to the same height. Again, since the memory image for weight changes rapidly, *S*'s judgment, in so far as it is based upon the image of the first weight, would be appreciably altered if the second weight were lifted at varying intervals after the first: the interval should, accordingly, be made as constant as possible, and fairly short, say not over 3 sec., and the arrangement of the weights must be such as to permit this procedure. Finally, in this test, since the suggestion hinges upon the visual perception of the block, *E* must be sure that *S* looks directly at each block as he lifts it.

Treatment of Data.—Following Gilbert, Scripture, and Seashore, the force of suggestion produced by the difference in size of the two standard blocks may be indicated by the difference in weight, in grams, between the two comparison blocks that are selected by *S* as the equivalents of the two standards.

The force of the size-weight illusion has been expressed by Scripture, on the basis of the more elaborate suggestion-blocks used by Seashore, in the form of a special law.[4]

RESULTS.—(1) *Dependence on age.* Sample results for normal children are those of Gilbert (Table 121) : it will be seen from them that the illusion is well developed at the age of 6 years, increases gradually till 9 years, and thence declines slowly with age.[5] Dresslar (6), however, whose method was undoubtedly less satisfactory, judged the effect of age to be indifferent for 7 years and above. Philippe and Clavière, who tested children from 3 years up, declare that the illusion decreases progressively below the age of 7, that it is obtained by

TABLE 121

Force of Suggestion (Gilbert)

AGE	6	7	8	9	10	11	12	13	14	15	16	17
NB	45	50	46	47	49	43	54	45	47	49	47	43
NG	47	45	46	47	42	48	49	58	53	51	39	41
F	42.0	45.0	47.5	50.0	43.5	40.0	40.5	38.0	34.5	35.0	34.5	27.0
P	36	37	27	36	23	22	15	8	7	12	6	5
MV	17.0	15.5	13.5	10.5	12.5	11.5	9.0	9.0	9.5	10.5	10.0	12.0
FB	43.5	43.5	45.0	50.0	40.0	38.5	38.0	37.0	31.0	33.0	32.0	25.0
FG	42.0	43.5	49.5	49.5	44.0	40.0	41.0	38.0	33.5	38.0	38.5	31.0

NB = number of boys
NG = number of girls
F = force of suggestion, in grams, for both sexes (median values)
P = per cent. of cases in which *F* exceeded 65 g., the limit used
MV = statistical mean variation
FB = force of suggestion, in grams for boys (median values)
FG = force of suggestion, in grams, for girls (median values)

[4]For the data from which this law is derived, see Scripture (19, p. 276f), also Seashore (15, pp. 3-14). For a striking demonstration of the force of the illusion, reference may be made to Wolfe's statements that "about one woman in 7 finds 1 g. of lead equal in weight to 60 g. of inflated paper bag," and not "one woman in 7 will find a gram of inflated paper bag half as heavy as a gram of lead" (21, p. 460).

[5]Gilbert's explanation is given in the following terms: "At 6 he has not yet learned to compare. As he learns gradually to judge a thing from more aspects than one, or in other words, learns to interpret one sense by another, the force of suggestion given by the eye to the muscle increases until at 9 he has come to the age of experience enough to see that things are not always what they seem. Consequently at this age he begins to correct misleading influences bearing upon him."

only a third of the children 3 to 6 years old, that it would perhaps disappear entirely if the test could be carried below the age of 3. They also found that the illusion is sometimes reversed in children of these ages.

(2) *Dependence on mental age (defective children).* In 1900 Demoor and Daniel (4) used the size-weight illusion, together with other tests, in an examination of 380 'abnormal' children, 6 to 15 years of age, in the city of Brussels. Ten of these children, all of them ranked as idiots or 'simple-minded,' either failed to get the normal illusion or had the illusion reversed. Three years later Claparède (3) obtained similar results with 18 mentally defective children at Geneva, and proposed that 'Demoor's sign' (failure to get the normal illusion) should be regarded as indicative of a diagnosis of medical (mental) retardation. In 1913 Doll (5) reported upon the examination at the Vineland Training School of 345 feeble-minded, of chronological ages 5 to 60, and mental ages 1 to 12 years. The results (Table 122 and Fig. 75) show that ability to perform the test at all (with or without getting the illusion) indicates a mental age of 4 years or over, while getting the normal illusion indicates a mental age of 7 years or over, since 84 per cent. succeed at 7 years and 100 per cent. at 8 years and above.

TABLE 122

Reactions of Feeble-Minded Children to the Size-Weight Illusion (Doll)

MENTAL AGE	NUMBER TESTED	COMPLETE FAILURE		NO ILLUSION		ILLUSION	
			%		%		%
1	35	34	97.1	1	2.9	0	0.0
2	37	28	75.7	3	8.1	4	10.8
3	38	17	44.8	8	21.6	12	31.6
4	32	5	15.6	14	43.7	13	40.7
5	35	5	14.3	9	25.7	21	60.0
6	36	2	5.6	12	33.3	22	61.2
7	45	3	6.7	4	8.9	38	84.4
8	41	0	0	0	0	41	100.0
9	25	0	0	0	0	25	100.0
10	12	0	0	0	0	12	100.0
11	4	0	0	0	0	4	100.0
12	5	0	0	0	0	5	100.0
Total	345	94		51		197	

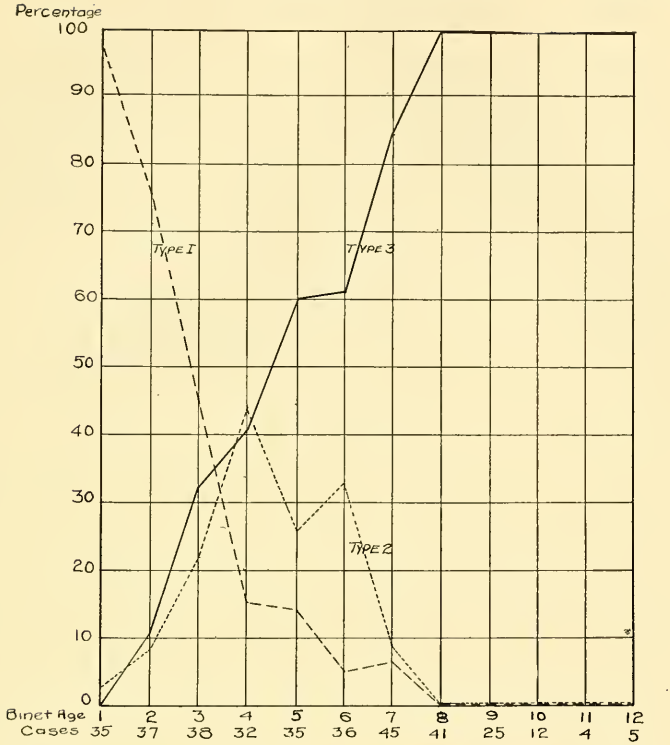

FIG. 75. SIZE-WEIGHT ILLUSION IN THE FEEBLE-MINDED. (Doll).

(3) *Dependence on sex.* The relation of sex to suggestion by the size-weight illusion has been differently stated by different investigators. Dresslar, for example, concludes that boys are more suggestible than girls. Wolfe, on the contrary, states that "men are less prone than women to illusions of weight," and that, in comparing wooden with lead weights, "the women overestimate the lead nearly twice as much as the men." Gilbert and Seashore find females more suggestible than males, but in nothing like the degree stated by Wolfe. Thus, inspection of his table shows that, according to Gilbert's method, after the age of 9, girls are, on the average, more influenced by the illusion than are boys. Seashore (16) tested 17 women and 28 men with two test-weights quite different in size, and found, similarly, that on the average the women showed the stronger illusion.

(4) The *relation of intelligence* to suggestibility among normal children has not been treated as carefully as the problem warrants. Gilbert made no correlations with intelligence. Dresslar concluded that bright children exhibit a stronger illusion, but Seashore (15) contends that Dresslar's method (arrangement in serial order) did not afford a real measure of the strength of the illusion.

(5) *Practise*, even if regular and persistent, does not dispel the illusion. It may, on the contrary, increase in amount (Hollingworth). If *S* be told the nature of the illusion, it still persists, though its intensity is thereby somewhat reduced (Seashore).

(6) If the method of procedure be modified, the strength of the illusion will be altered.

The more important of the relations thus revealed are the following:[6]

(*a*) "The illusion of weight dependent on size is greatest when size is estimated mainly by muscle-sense, and the weights have not previously been seen." Fourche says it is then three times as strong.

(*b*) "The illusion is more fluctuating and on the whole not quite so strong when size is estimated by the area of pressure in the flat palm, including a memory of the third dimension."

(*c*) "In these variations, the illusion is weakest when size is estimated by direct sight."

(*d*) "When size is estimated by the combined effect of all the spatial senses, the illusion is weaker than when depending on muscle-sense or touch, and stronger than when dependent on sight alone."

(*e*) The illusion is weaker when the blocks are viewed in indirect vision, and still weaker when judged by visual memory.

(*f*) A knowledge, or supposed knowledge, of the material of which weights are made may affect the estimate of their weight.

(*g*) The illusion does not necessarily vary directly with the volume of the compared weights, but depends in part upon the manner in which the difference in volume is brought about.

(*h*) The illusion obtains among the blind, where it follows the same general law as for the seeing, though it is not so strong, either for lifted or merely 'touched' weights, as for the seeing under the same conditions (Rice).

NOTES.—The outcome of any test of weight-comparison is somewhat affected by the tendency felt by all *S*'s, though differing in degree between different individuals and in the same individual at different times, to overestimate the second of two lifted weights.

[6] See, especially, Seashore (15).

If blocks of different material, *e. g.,* cork and lead, or wood and iron, be constructed in such a manner as to have the same dimensions and the same weight, the knowledge of the actual differences in the weight of the two materials produces an illusion similar to the size-weight illusion. Seashore (16) tested school children with this *material-weight illusion,* and found that the overestimation of the metal blocks amounted to from 7 to 11 grams (or from 13 to 20 per cent. of their actual weight, 55 g.). For this illusion, it is of interest to note, there was found virtually no variation with age, sex, or intellectual ability.

REFERENCES

(1) F. E. Bolton, A contribution to the study of illusions, etc. *AmJPs*, 9: 1898, 167-182, especially 167-178.

(2) A. Charpentier, Analyse expérimentale de quelques éléments de la sensation de poids. *Archives de physiologie normale et pathologique*, 5th ser., 3: 1891, 122-135, especially 126ff.

(3) E. Claparède, L'illusion de poids chez les anormaux et le 'signe de Demoor.' *ArPs (f)*, 2: 1903, 22-32.

(4) Demoor et Daniel, Les enfants anormaux à Bruxelles. *AnPs*, 7: 1900 (1901), 296-313, especially 307-8.

(5) E. A. Doll, The Demoor size-weight illusion. *TrSc*, 9: 1913, 145-149.

(6) F. B. Dresslar, Studies in the psychology of touch. *AmJPs*, 6: 1894, 313-368, especially 343-360.

(7) Th. Flournoy, De l'influence de la perception visuelle des corps sur leur poids apparent. *AnPs*, 1: 1894 (1895), 198-208.

(8) J. A. Fourche, L'illusion de poids chez l'homme normal et le tabétique. Nancy, 1911.

(9) J. A. Gilbert, Researches on the mental and physical development of school children. *StdYalePsLab*, 2: 1894, 40-100, especially 43-5, and 59-63.

(9a) H. L. Hollingworth, The influence of caffein on mental and motor efficiency. *ColumbiaContPhPs* 20, and *ArPs(e)*, No. 22: 1912, pp. 166, especially p. 20.

(10) H. N. Loomis, Reactions to equal weights of unequal size. *Std YalePsLab*, n. s. 1: No. 2, June, 1907, 334-348. (Same as *PsMon*, 8: No. 3, whole No. 34.)

(11) L. Martin und G. E. Müller, Zur Analyse der Unterschiedsempfindlichkeit. Leipzig, 1899. Pp. 233.

(12) G. E. Müller und F. Schumann, Ueber die psychologischen Grundlagen der Vergleichung gehobener Gewichte. *ArGsPhg*, 45: 1889, 37-112.

(13) J. Phillipe et J. Clavière, Sur une illusion musculaire. *RPhF*, 40: 1895, 672-682.

(14) J. F. Rice, The size-weight illusion among the blind. *StdYalePs Lab*, 5: 1897, 81-87.

(15) C. E. Seashore, Measurements of illusions and hallucinations in normal life. *StdYalePsLab*, 3: 1895, 1-67, especially 1-29.
(16) C. E. Seashore, The material-weight illusion. *UnIowaStdPs*, 2: 1899, 36-46.
(17) E. W. Scripture, Remarks on Dr. Gilbert's article. *StdYalePs Lab*, 2: 1894, 101-4.
(18) E. W. Scripture, The law of size-weight suggestion. *Sci*, n. s. 5: February 5, 1896, 227.
(19) E. W. Scripture, The new psychology. London, 1897. See ch. xix.
(20) J. van Biervliet, La mesure des illusions de poids. *AnPs*, 2: 1895, 79-86.
(21) H. K. Wolfe, Some effects of size on judgments of weight. *PsR*, 5: 1898, 25-54.

TEST 41

Suggestion by progressive weights.—This test, like that which follows it, is one of several devised by Binet for the purpose of securing a quantitative measure of the degree of suggestibility of children or adults when the suggestion is 'depersonalized,' in the sense that it is derived by *S* himself from the objective conditions of the experiment, rather than from the attitude, tone, instructions, or personality of *E*. The principle embodied in this test is, in other words, the arousal, by auto-suggestion, of a "directive idea," or the rapid development of an attitude of expectation. Suggestibility is measured, at least approximately, by the ease with which this suggestion, or habit, of judgment, is aroused and by the persistence that it displays under conditions which tend gradually to counteract it.

Materials.—A set of 15 weights, of identical size and appearance, numbered conspicuously from 1 to 15. The first four weigh 20, 40, 60, and 80 grams, respectively; the remaining 11 weigh 100 grams each. Table of such a height that *S* can stand in front of it and lift the weights readily. A thick gray or black cloth.

Preliminaries.—Spread the cloth over the table. Place the 15 weights in a line as numbered, with the lightest on the left and the 11 heaviest on the right, and with about 2 cm. between each weight. No. 1 is then at the left, No. 15 at the right of the row.

Method.—Give *S* the following instructions: "Here is a series of weights, 15 of them. I want you to lift them, one after the other, like this. [Illustrate by taking a weight between

thumb and finger and lifting some 10 cm. from the table.] As you lift each weight, I want you to tell me whether it is heavier, lighter, or the same as the one just before it. All you have to say is either 'heavier,' or 'lighter,' or 'the same.' Remember you are to compare each weight with the one you lifted just before. For instance, when you lift the 8th, you are to say whether it is heavier, lighter, or the same as the 7th. Here is the first weight, number one, at the left end of the row."

Watch *S* to see that he follows these instructions, particularly that he lifts the weights successively, without relifting earlier ones. Record his judgments verbatim; be careful, also, to note any secondary evidences that might throw light on his judgments, *e. g.*, attitudes or expressions of hesitancy, assurance, surprise, embarrassment, cautiousness, etc.

VARIATIONS OF METHOD.—(1) In the second method followed by Binet, *S* is instructed to lift, in each trial, the preceding weight as well as the one that is being judged, *e. g.*, he lifts the 8th, then the 7th, then the 8th again: next the 9th, then the 8th, then the 9th again, etc. The lifting is all done, as before, with the one hand.

(2) In the third method followed by Binet, *S* is asked to estimate the first weight lifted. He usually gives too small an estimate. He is then told that its weight is 20 grams (about 0.7 ounce). The series is now compared, using either of the methods of lifting above described, according to *S*'s preference,[1] but *S* is required to estimate or guess the heaviness of each weight, basing his judgment, of course, merely on the knowledge that the first weight is 20 grams.

TREATMENT OF DATA.—From the tabulated results, *E* may easily determine in how many cases the objective progression of the first 5 weights was correctly noted. For a measure of suggestibility, *E* must take the number of times 'heavier' is judged in the last 10 judgments (when 'same' is the correct judgment). This measure is admittedly somewhat crude, but

[1] It would, obviously, be better to prescribe either the one or the other method for all *S*'s. The first method has the merit of taking less time, and it is the method that is for the most part naturally adopted by younger *S*'s.

it affords a fairly reliable index for determining the relative order of rank of a group of *S*'s. Thus, an *S* that judges 'heavier' 10 times is unquestionably more suggestible than one who answers 'heavier' but 5 times, though not necessarily twice as suggestible.

If all three methods are employed, *E* may determine *S*'s suggestibility by adding the number of false 'heavier' judgments in all three tests. In the third method, the quantitative estimate given by *S* for the 15th weight (or the maximal estimate for weights 6 to 15) might be taken, in comparison with his estimate of the 5th weight, as an index of suggestibility, but this method is not regarded by Binet as so reliable as the one already described.

Results.—(1) The general outcome of the test as conducted by the first, or standard, method is indicated in Table 123, which embodies the results obtained by Binet upon 24 elementary-school children, aged 8-10 years.

(2) It is evident that, in children of this age (8-10), not all judge correctly the actual objective increase in the *first five weights*. Since the differences are supraliminal, the exceptions

TABLE 123

The Progressive-Weight Suggestion. 24 Cases (Binet)

NO. OF WEIGHT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Actual weight	20	40	60	80	100	100	100	100	100	100	100	100	100	100	100
Times estimated +		24	19	19	23	13	18	18	18	12	19	19	17	15	12
Times estimated −		0	1	1	0	9	1	4	1	7	2	2	3	3	5
Times estimated =		0	4	4	1	2	5	2	5	5	3	3	4	6	7

must be ascribed to faulty attention, though, possibly, the fact that the weights are of equal size may have clouded the direct perception of weight by lifting.

(3) In general, the suggestion is still working, though less powerfully, at the 15th trial: in other words, it has persisted, for most *S*'s, through the successive lifting of 10 equal weights.

(4) There is a marked drop in the judgment 'heavier' at *the 6th weight, i. e.,* at the first 'trick' weight—a drop which is,

obviously, due to a 'disappointed suggestion,' analogous to that which conditions the size-weight illusion of the preceding test. In the present instance, *S* is, in most cases at least, prepared to find the 6th weight heavier than the 5th: he puts forth more effort; the weight rises with unexpected ease, and is, therefore, often judged 'lighter.' If, however, *S* is more influenced by his expectation of 'heavier' than by the unexpected lightness of the weight, he still judges 'heavier,' or he may, from the conflict of these two tendencies, judge 'equal.'

(5) *Practise* has very little effect upon the suggestibility of *S*'s: at least Binet found that, when 12 older children (16 years) repeated the test by the first method five times in immediate succession, there was no alteration in the average number of times that suggestion appeared (the average number of suggestions in the five trials was 5.1, 4.9, 5.4, 5.0 and 5.5, respectively).

(6) Binet's tentative experiments indicated that *age* apparently has less effect upon suggestion by progressive weights than upon suggestion by progressive lines (see the following test). In trials by the first method, 12 children aged 16 years responded, on the average, with 5.1 suggestions, whereas 24 children aged 8-10 years, responded, on the average, with 6.75 suggestions. The later experiments conducted under Binet's direction by Giroud show a distinct lessening of suggestibility after 9 years when only those *S*'s are considered who made no errors in the first four judgments (objective increase of weight), as is indicated in the last column of Table 124.

TABLE 124

Averages for Progressive-Weight Suggestion by Age (Giroud)

AGE	HEAVIER JUDGMENTS	CONSECUTIVE HEAVIER JUDGMENTS	EQUAL JUDGMENTS	HEAVIER JUDGMENTS OF S'S MAKING NO ERROR WITH OBJECTIVE INCREASES
7	6.5	5.5	2.3	8.4
8	6.6	5.3	1.6	7.0
9	5.2	2.4	2.0	8.0
10	7.0	4.8	1.2	4.5
12	5.9	3.6	5.0	5.0

(7) According to Binet, comparison of the results of this test with *other tests of suggestibility*, especially the line-test, indicates a fair degree of correlation, so that, while the sense-department under examination may in part determine the extent of suggestion, very suggestible *S*'s may be expected to prove noticeably suggestible in all tests. On the other hand, tests undertaken in the Educational Laboratory at Cornell University[2] do not confirm Binet's statement, and lead one to believe that Scott's conclusions (Test 44) are correct, when he asserts that there is no such thing as general suggestibility.

(8) Procedure by the *second method* (compulsory lifting of the antecedent weight) makes the real progression (1st five weights) more uniformly evident, but reduces the illusory progression.

(9) Procedure by the *third method* (estimates of each weight) produces less suggestion than the first, but more than the second method. Inspection of the estimated weights (grams) show (*a*) that *S*'s have a decided preference for the use of numbers terminating in 0 or 5, (*b*) that no one of the 24 *S*'s overestimated the 5th weight (100 g.), but that they commonly greatly underestimated it (30 to 50 g.), and (*c*) that those *S*'s that showed the greater number of suggestions also gave, on the average, the largest quantitative estimations for the illusory increments. The correlation of suggestibility under these two methods of treatment (first and third) was found by Okabe and Whipple to be 0.53.

REFERENCES

(1) A. Binet. La suggestibilité. Paris, 1900. Ch. iv. (pp. 161-208).
(2) A. Giroud. La suggestibilité chez des enfants d'école de sept à douze ans. *AnPs*, 18: 1912, 362-388.

TEST 42

Suggestion by progressive lines.—The purpose and general plan of this test are the same as in the preceding test of sugges-

[2]These tests, which were conducted by T. Okabe, under the author's directions, included all the suggestibility tests of Binet, together with the warmth tests (No. 44). The results of their application to 29 *S*'s indicate almost total lack of correlation of suggestibility in the several tests.

tion by progressive weights, and the details are again derived from the work of Binet.

APPARATUS.—A sheet of cross-section paper, ruled in millimeter squares. Kymograph drum, with kymograph or some form of supporting stand. Cardboard. Strip of white paper, 15×15 cm. Drawing materials.

PRELIMINARIES.—Arrange the kymograph drum so that it may lie horizontally and be revolved freely by hand. It may conveniently be left in the kymograph with the driving 'step' loosened, or be placed in the smoking stand. Across the strip of white paper, draw with a ruling pen 20 parallel, straight, black lines, 2 cm. apart and each 1 mm. wide. The lines must begin at varying distances from the left-hand margin: the first four are to be 12, 24, 36, and 48 mm. long, respectively; the remaining 16 are to be each 60 mm. long. Support the sheet of cardboard vertically in front of and close to the kymograph drum, and cut a horizontal slit 1×12 cm. through the cardboard in such a position as to expose the ruled lines, one by one, as they are turned past the slit.[1]

METHOD.—Seat *S* 50 cm. from the screen and provide him with a sheet of cross-section paper. The instructions should take the following form: "I want to try a test to see how good your 'eye' is. I'll show you a line, say an inch or two long, and I want you to reproduce it right afterwards from memory. Some persons make bad mistakes; they make a line 2 inches long when I show them one 3 inches long; others make one 4 or 5 inches long. Let's see how well you can do. I shall show the line to you through this slit. Take just one look at it, then make a mark on this paper [cross-section paper] just the distance from this edge [left-hand margin] that the line is long. When that is done, I shall show you the second line, then the

[1] In default of the kymograph, the strip of ruled lines may be laid flat upon the table and exposed through a 1 × 12 cm. slit cut in the center of a sheet of cardboard 55 cm. square.

Or, the test-lines may be drawn as sections of radii upon a cardboard disc which is supported vertically just behind the screen and rotated to bring them into view successively.

third, and so on. Make the marks for the second on the line below the first, the third on the next line, and so on."[2]

E then turns the drum to bring the first, or shortest, line into view. As soon as *S* turns his attention to the recording of his estimate on the paper, the drum is moved forward slightly to conceal the line, so that further comparison is impossible. As soon as *S* has placed his mark, then, and not before, the next line is exposed. This precaution serves to maintain the impression that a new, and hence probably a longer line is exposed. Slow *S*'s may need to be hurried; too quick ones may need to be checked, so that the interval between successive exposures shall be approximately 7 sec. To keep *S*'s attention alive, *E* may accompany the exposures with non-suggestive remarks, *e. g.*, "Here is the second line." "Here is the third," etc.

If *S* has ceased to respond to the suggestion of progressive augmentation at the 20th exposure, the test ends at that point: if not, *E* should, without *S*'s knowledge, bring the drum back to the 5th line, and continue the exposures of the series of 60 mm. lines as before, until *S* does cease to respond to the suggestion.

E should note and record any significant features in *S*'s manner, *e. g.*, signs of embarrassment, hesitancy, automatic response, etc.

When the test is completed, and provided no further tests of suggestibility are to be undertaken at the time, *E* will find it advantageous to quiz *S* with regard to his attitude toward the test. This interrogation must be very tactfully conducted. *E* may, for example, begin by saying: "Are you entirely satisfied with what you have done"? If *S* answers in the affirmative, let *E* continue with such inquiries as: "Do you think you have made any mistakes"? "Did you make any lines too short or too long"? "At what moment did you notice that your lines were too long"? "Why didn't you make them shorter"? etc.

[2]These directions should be followed with some care. In tests of suggestion, the slightest change in the setting of the test, or in the manner or content of the instructions, may materially affect *S*'s attitude toward the experiment. The object is to convey the idea of a straightforward test of accuracy of line-reproduction, and to avoid arousing any suspicion of snares or tricks.

If *S* confesses that he made some mistakes, let him take his record-sheet and make the changes that he thinks ought to be made to produce a correct record, using small circles for his corrections to avoid confusion with his first estimates.

VARIATIONS OF METHOD.—*E* may, if desired, adopt the arrangement first used by Binet, according to which there are 12 successive stimulus-lines, all of which begin at the same distance from the left-hand margin, and which have the following lengths: 12, 24, 36, 48, 60, *60*, 72, *72*, 84, *84*, 96, *96* mm. It is evident that numbers 6, 8, 10, and 12 constitute four 'trap-lines,' since the arrangement suggests progressive augmentation, whereas each of these four lines is equal to that which immediately precedes it.

TREATMENT OF DATA.—(1) For a measure of suggestibility, *E* may take the number of lines out of the last 15 lines that are drawn longer than the 5th line was drawn.

(2) A coefficient of suggestibility may also be calculated, following Binet's method, by the formula

$$x : 100 = max.\ L : 5th\ L,$$

in which

x = the required coefficient,

max. L = the length of the maximal line recorded by *S*,

5th L = the length of the 5th line as recorded by *S*.

Absence of suggestibility is, then, indicated by a coefficient of 100: presence of suggestibility by a coefficient of over 100.

(3) When the variant method is used, the degree of suggestibility may be determined roughly in terms of the number of 'traps' in which *S* is 'caught,' or more exactly, by the formula

$$x : 100 = c : r,$$

in which

x = the required coefficient,

c = the average recorded increment of the four trap lines,

r = the average recorded increment of the four lines immediately preceding the four trap-lines.

RESULTS.—(1) In his examination of pupils in the elementary schools, aged 8-10 years, Binet found that *the coefficient of suggestibility* ranged from 109 to 625. In 16 of 42 pupils, the coeffi-

cient was 200 or over, *i. e.*, the maximal line was double or more than double the 5th line.[3]

With the variant form of test, Binet found the coefficient lying between 7.6 and 120. No one of 45 children avoided all four 'traps,' and 36 children avoided none of them. Occasionally, the trap-line, presumably on account of the contrast between the stimulus and the child's expectation, was actually recorded as shorter than the preceding line. *S*'s whose coefficient in this form of test is 100, *i. e.*, whose average increment for the trap-lines is the same as for the objectively progressive lines, are termed 'automatic.'

(2) The point at which *maximal suggestion* is registered is commonly between the 19th and the 25th line, but may lie anywhere between the 7th and the 36th (this being the limit tested by Binet).

(3) Inspection of the records of individual pupils shows that in some cases the force of suggestion was steady and persistent, while in others it reached a maximum, and then declined.

(4) *Extremely suggestible S*'s may make their 'estimate' of the line without even looking at it when exposed; their minds are so completely dominated by the suggestion of uniform augmentation that they do not trouble to observe the stimulus.

(5) The degree of suggestion induced by this test declines markedly with *age:* Binet found, for instance, that the coefficients of suggestibility, in the case of 12 pupils whose age averaged 16 years, ranged only from 103 to 146. Binet's work was continued by Giroud, who tested 38 children and obtained for averages, on the basis of a possible score of 15, the following averages for the number of times any line beyond the 5th exceeded the length assigned to the line just before it:

YEARS	7	8	9	10	12
Cases tested	10	10	5	5	8
Score	10.7	8.2	4.2	4.8	1.0

That it is difficult to induce the suggestion with adults was shown in scattered tests made by the author upon college students. On the other hand, Chojecki, who tested 30 men and 30

[3]For detailed records of number of individual cases, consult Binet, 124ff.

women students of the University of Geneva, had better success, and found, indeed, 36 persons (22 men and 14 women) susceptible of suggestion by this method.

(6) In either form of test, the *1st line* is apt to be over-estimated. The *5th line* is almost invariably underestimated. Generally speaking, this underestimation is less pronounced in those *S*'s that prove least suggestible.

(7) In many instances, the records bear witness to a struggle between the directive idea of progressive increments and the impressions which are actually received from the lines as they are exposed. Especially characteristic is the appearance of a number of estimates in which the directive idea is effective, followed by a sudden reduction in estimation, which is again followed by another series of progressive increments. In other words, the idea of progression is operative until a point is reached when the recorded length is manifestly too long. *S* makes, then, a more or less marked correction, but does not, curiously, relinquish the notion of progression, and this again becomes manifest.

(8) The corrections made by young *S*'s during the inquiry that follows the test cannot, of course, be taken as exact indications of the extent of the suggestion or of their consciousness of error. It will be found that many *S*'s are conscious that they have made the lines too long; some can also explain why they made them too long; but it is rare that any one gives a satisfactory explanation of why he continued to make them too long, after he realized that he had been overestimating.

(9) *Correlations.* Tests of school children and of adults by Okabe and Whipple afforded the following correlations (foot-rule method): Suggestibility for progressive lines (number of 'yields') and suggestibility for progressive lines (maximal divided by the 5th line) 0.38; correlation, by either treatment, with contradictory suggestion, (Test 43) about 0.25, with directive suggestion (Test 43) about 0.20, with suggestion for warmth 0.17, with the size-weight illusion (Test 40) 0.10 by the first, and —0.14 by the second method of computing suggestion for progressive lines.

REFERENCES

(1) A. Binet, La suggestibilité. Paris, 1900. Pp. 83-160.
(2) A. Chojecki, Contribution a l'étude de la suggestibilité. *ArPs(f)*, 11: 1911, 182-186.
(3) A. Giroud, La suggestibilité chez des enfants d'école de sept à douze ans. *AnPs*, 18: 1912, 362-388.

TEST 43

Suggestion of line-lengths by personal influence.—In the three preceding tests suggestion is produced by the objective conditions of the test: in everyday life, however, suggestion is often produced by personal influence, by authoritative statement or command, or merely by what Binet terms 'moral influence.' Two forms of line-test have been utilized by Binet to study this variety of personal suggestion: the first he terms 'contradictory suggestion,' the second 'directive suggestion' (*suggestion directrice*): in the former *E* makes certain statements that are intended to interrupt or modify a judgment that *S* has just made; in the latter, statements that are intended to control or influence a judgment that *S* is just about to make.

A. CONTRADICTORY SUGGESTION

MATERIALS.—Drawing materials. A sheet of cardboard upon which are drawn in ink 24 parallel, straight, black lines, ranging in length from 12 to 104 mm., by increments of 4 mm. The lines all begin at the same distance from the left-hand margin, are 7 mm. apart, and are numbered in order of their length, from 1 to 24. Three rectangular pieces of cardboard, about 12x20 cm., on each of which is drawn a single straight line. These three stimulus-lines correspond to numbers 6, 12, and 18 of the 24 comparison-lines, and are, accordingly, 32, 56, and 80 mm. long, respectively.

METHOD.—Show *S* the card of comparison-lines, and explain their numbering. Replace this by the first stimulus-line (32 mm.), saying: "Look carefully at this line." After 4 sec., remove the stimulus-card, present the comparison-card, and say: "Tell me the number of the line that is just the length of the one I showed you." At the moment that *S* gives his judgment, *E* says: "Are you sure? Isn't it the —th"?—indicating always the next longer line. If *S* answers "No," *E* repeats the question

in exactly the same form. If *S* still answers "No," the attempt to produce suggestion is suspended, and the case is recorded as one 'resistance.' The second and the third stimulus-lines are presented and the same procedure is followed in each case. If, in any of the trials *S* answers "Yes," *E* then inquires: "Isn't it this one"?—indicating the next longer line, and this inquiry is carried on from line to line until *S* has twice resisted the suggestion, *i. e.*, has twice answered "No" to the same question.[1]

VARIATIONS OF METHOD.—For many *S*'s, particularly for adults, more success will attend the use of a second method tried by Binet in preliminary tests, viz.: the introduction of an interval of 12 sec. between removal of the stimulus-line and presentation of the comparison-card.

TREATMENT OF DATA.—Following Binet, *S*'s suggestibility may be rated in terms of the total number of 'advances' in lines that he makes, under inquiry, in all three trials. Thus, if he 'yields' two lines the first time, three the second, and none the third, his suggestibility is rated as 5.

RESULTS.—(1) Children tend to select for their first line one that is shorter than the stimulus-line.[2]

(2) Of 25 children, aged 8-10 years, Binet found 6 who resisted suggestion completely, 6 who 'yielded' once, 5 twice, 2 three times, 2 four times, and one each six, seven, and more than seven times.

(3) Preliminary experiments conducted by Binet and Henri upon 240 pupils, with some slight changes in method (particularly, giving an opportunity both for direct comparison and for selection by memory after a 12 sec. interval), yielded the results (2: p. 343) indicated in Table 125.

Here it is evident that *E*'s suggestion is less effective when *S* can make direct comparison of the lines, and that suggestibility,

---

[1]Once more it should be said that it is highly important to follow the same form of inquiry, to use the same tone, the same attitude, in every question for every *S*, since the suggestion which we seek to measure is conditioned by the character of the inquiries.

[2]*E* is almost always, therefore, in a position to demonstrate to *S*, if need be, after the test, that his suggestion would have been a sound one to follow.

TABLE 125

Percentage of 'Yields' to Contradictory Suggestion (Binet and Henri)

AVERAGE AGE	MEMORY TEST	COMPARISON TEST	MEAN
7–9	89	74	81.5
9–11	80	73	76.5
11–13	54	48	51.0

under either direct comparison or comparison from memory, declines with *age*.

(4) *S*'s who have selected the correct line are less apt to change their designation under suggestion than are *S*'s who have selected the wrong line: thus Binet and Henri found that 56 per cent. changed their selection when it was actually right, but 88 per cent. when it was wrong. Moreover, of the latter, 81 per cent. made the change in the proper direction.

B. DIRECTIVE SUGGESTION

Apparatus.—As in Test 42, save that only the 60 mm. lines are used.

Method.—Seat *S* 50 cm. from the cardboard screen and provide him with a sheet of cross-section paper. Instruct him as follows: "I'm going to show you a number of lines. You will see them appear through this slit, one at a time. When I show you a line, take a good look at it; then make a mark on this paper at just the distance from this edge [left-hand] that the line is long. When that is done, I shall show you the second, then the third, and so on. You will make the mark for the length of the second line on the second line of your paper, for the third on the next line, and so on."

E now displays the 5th, *i. e.*, the first 60 mm. line of the series, with the remark: "Here is the first one." When *S* is ready for the second line, *i. e.*, 7-10 sec. later, *E* remarks, as he exposes it: "Here is a longer one." When the third is exposed, he remarks "Here is a shorter one;" and he continues to use these remarks, alternately, at the moment of exposure of each line, until 15 lines have been exposed, the first without suggestion, the remainder coupled with 14 suggestions—7 of shorter, 7 of longer.

These suggestions must be given just before the line is exposed, in a quiet tone, without looking at *S*. *S* should see the disc turn and the new line appear at the moment that he receives the suggestion.

If desired, *S* may be questioned afterward, as indicated in Test 42, with regard to his attitude toward the suggestions.

TREATMENT OF DATA.—When *S* accepts the suggestion, record a 'plus' case; when he resists the suggestion, either by making the length equal to that of the preceding line, or by altering the length in a direction contrary to the intent of the suggestion, record a 'minus' case. The number of the plus cases may serve as an index of *S*'s suggestibility. Record should also be kept of the extent of modification (in mm.) made by *S* in each trial.

RESULTS.—(1) The verbal directive suggestion used in this test is more potent, at least for children 8-10 years old, than the auto-suggestion induced in Test 42. Sixteen of 23 pupils tested by Binet submitted completely to the suggestion,[3] and no one resisted every suggestion.

(2) The suggestion is, in general, stronger at the outset than toward the end of the series, as is indicated by the fact that the extent of modification of line-length decreases, and the number of complete resistances increases, as the series progresses.

(3) Verbal suggestion is commonly more effective in producing augmentation than in producing reduction in line-length, in the proportion of about 5 to 4.

(4) There are marked *individual differences* in the suggestibility of school children under the conditions of this test. Binet found that in 18 trials the number of resistances to suggestion ranged from 0 to 14. (See Binet, 1, pp. 228-9, for a detailed table.)

(5) The *first line* is practically invariably underestimated.

(6) Tests upon 10 children, whose average *age* was 17 years, showed less suggestibility than in the case of younger children; still, 4 of the 17 accepted every suggestion, and 3 others resisted suggestion only once. The average extent of modification produced by suggestion is, however, less than in the case of

[3] This statement is made in the text, but does not appear to be borne out by Binet's table (1, pp. 228-9).

younger *S*'s. Again, the extent of modification is practically constant throughout in the series with the older *S*'s, but large at first and then progressively less in the series with the younger *S*'s.

NOTES.—The experiments of Bell, Brand and Jones, in which the estimates or judgments of spatial magnitudes or extents were subjected to verbal suggestions, such as "make high," "make low," "you are now able," "you are now unable," etc., cannot be directly compared with the work of Binet, because in all of them the *S*'s were well aware of the intentional and artificial character of the suggestions and were instructed to avoid voluntary resistance to them. In general, the suggestions in these experiments had some effect upon the work of the *S*'s, but not upon all of them, nor always in the direction in which they were supposed to influence the outcome.

REFERENCES

(1) A. Binet, La suggestibilité. Paris, 1900, especially 219-243.
(2) A. Binet et V. Henri, De la suggestibilité naturelle. *RPhF*, 38: 1894, 337-347.
(3) J. C. Bell, The effect of suggestion upon the reproduction of triangles and of point distances. *AmJPs*, 19: 1908, 504-598.
(4) J. E. Brand, The effect of verbal suggestion upon the estimation of linear magnitudes. *PsRev*, 12: 1905, 41-49.
(5) Grace M. Jones, Experiments on the reproduction of distance as influenced by suggestions of ability and inability. *PsRev*, 17: 1910, 269-278.

TEST 44

Suggestion by illusion of warmth.—In measuring either discriminative or liminal sensitivity, difficulty is not infrequently caused by the interference of auto-suggestion (see various tests of Chapter VI). In the immediately preceding tests (Nos. 40 to 43), a process of discrimination (of weights and line-lengths) was, accordingly, made the basis for testing suggestibility. In the present test, a (supposed) measurement of liminal sensitivity is made the basis for testing suggestibility. The plan is to arrange experimental conditions in such a way as to suggest warmth, when no warmth is present.

This idea seems to have originated in the Yale laboratory, when Seashore (5), in 1895, worked out a proposal made two years earlier by Scripture (4). Small's varied tests of sug-

gestibility (6), which appeared in the following year, embodied two very simple 'heat' tests. More recently, Guidi in 1908, Scott in 1910, and Chojecki in 1911, have reported tests of suggestibility to warmth, the former with a simple 'warmth box,' the latter with apparatus somewhat similar to the original device of Seashore. Four methods are described herewith; the resistance-wire method of Seashore and Scott, the heated box method of Guidi, and the two simple methods employed by Small.

A. ILLUSORY WARMTH—RESISTANCE-WIRE METHOD

APPARATUS.—Stop-watch. Special warmth-tester.

The warmth-tester consists of a wooden box, open at the end facing *E*, and provided, on the top, with porcelain sockets for four electric lamps, wired in multiple, and with a snap switch by which the current (105-110 volt, D. C.) may be turned on or off. The wiring is purposely left visible, and leads conspicuously from the lamps to a coil of No. 24 German-silver wire, 1 m. long, which is wound, without covering, about a flat piece of hard rubber, 3 x 10 cm. This resistance coil is fastened to the front of the box, in such a manner that it may be easily reached by *S*, without exposing his fingers to the warmth of the lamps on the top of the box. A concealed circuit leads to a noiseless switch, underneath the box, which can be operated by *E* without *S*'s knowledge. By means of this switch, *E* may shunt the current through the coil, or cut the coil out entirely, without affecting the illumination of the lamps.[1]

PRELIMINARIES.—Find an arrangement of lamps such that, when the current passes through the coil, warmth becomes perceptible in 8 to 10 sec. Four 25-watt tungsten lamps generally prove satisfactory. If necessary, use one or more 40-watt lamps.

METHOD.—Give *S* the following instructions: "I want to test your ability to perceive warmth. Hold this coil of wire gently between your thumb and two fingers, like this [illustrating]. You will see that the coil is connected with these electric lamps, so that, when I light them, a current of electricity can flow through the coil and warm it—it is made of German-silver wire, and offers a slight resistance to the current. There is nothing at all to be afraid of. You can't feel any shock from

[1] In default of a 110-volt circuit, a resistance-wire apparatus may be contrived with a battery, after the plan described by Seashore, though the absence of the illuminated lamps alters the experimental conditions.

the current, nothing but a slight warmth. Watch carefully, and, the moment that you feel warmth, say 'now.'"

Without attracting *S*'s attention, close the secret coil-switch, so that no current passes through the coil. After a preliminary 'ready,' snap the lamp-switch rather ostentatiously; start the stop-watch at the same instant, and lean forward in an attitude of expectancy, keeping one hand on the lamp-switch, as if awaiting *S*'s 'now.' Snap the lamps off as soon as the 'now' is spoken. Record the time. Feel of the coil, or solicitously blow upon it, as if to cool it. Repeat the test 5 times with each hand, alternately.

If *S*, at any trial, fails to get the illusion of warmth within 60 sec., open the coil switch (without *S*'s knowledge), so that warmth is actually felt, but record the trial as one 'resistance,' or failure.

Variations of Method.—Following the plan of Seashore and of Scott, tell *S* that 20 trials will be made. Give a preliminary series of 5 trials with each hand, with objective warmth from the start, in each trial. Without interruption, continue with an equal number of trials in which the coil is not warmed unless *S* fails to report warmth within a period of some 10 sec. longer than the average time at which he had reported warmth in the first 10 trials.

Treatment of Data.—In either method, suggestibility is measured by the absolute or relative number of trials (without objective warmth) in which *S* reports warmth.

S may also be rated in terms of the quickness (number of seconds) with which the illusion is reported.

B. Illusory warmth—Guidi's method

Apparatus.—Stop-watch. Matches. Alcohol lamp, fitted with hinged extinguishing cap. Cubical wooden box, with a chimney-like metal top, a circular hole in the front face, and a hinged door in the back face (Fig. 76).

Method.—*E*'s instructions are analogous to those in the resistance-wire method. "I want to test your ability to perceive warmth. I want you to thrust your forefinger into this box through the hole in front. I shall put this lamp into the box.

It won't burn you at all. Just watch very carefully, and say 'now' the moment that you notice any warmth in the box." *E* then lights the alcohol lamp, opens the door of the box, sets in the lamps extinguishing the flame as he does so, starts the watch, closes the door, and expectantly awaits *S*'s judgments.[2]

MATERIALS.—Alcohol lamp. A pin thrust through the rubber tip of a pencil, or through a small bit of soft wood. Toothpick, or other bit of wood with a blunt point. Matches. Piece of cardboard, about 15x15 cm. Blindfold.

METHOD.—(1) Let *S* see the lighted lamp and the pin in its holder. Instruct him as follows: "I am going to warm this pin in this flame, then touch it to the back of your hand to see if you can notice the warmth it makes. Don't be afraid of being burned, as it will not be hot enough for that, and I shall try it on my own hand first. Say 'now' when you feel its warmth." Blindfold *S* carefully. Go through the operation of heating the pin; say 'ready,' but do not touch *S*'s hand at all. If *S* reports warmth, ask him to describe the feeling: if he does not report warmth, repeat the test, but touch him on the back of the hand with the pointed piece of wood, to see if the contact is reported as 'warm' or 'hot.'

(2) Light a match and move it around about 1 cm. above the back of *S*'s hand. Call his attention to the 'waves of heat' that he feels. Blindfold him carefully. Ask him to see if he can detect the heat waves every time. Strike a match, and move it about over his hand, but hold the cardboard between the match and the hand. Repeat several times with either hand. Note the number of times the suggestion is 'accepted,' and any indications of the readiness or degree of suggestibility.

RESULTS FOR ALL METHODS.—(1) In general, the results of the warmth-illusion test appear to be conditioned primarily by the success of the investigator in creating a proper atmos-

[2]Guidi's method deviated somewhat from the above, in that *S* was instructed to push his finger slowly into the box against a metal disc, and degree of suggestibility was measured by the extent to which the finger had been inserted when warmth was reported. This procedure presents difficulty in governing the rate of movement, and has, so far as the author's experience goes, no advantage over the procedure that has been recommended.

phere of suggestibility, rather than upon the particular apparatus employed. Thus, Seashore met with amazing success. Of his 8 college students, only 3 resisted at all, and these but once or twice each, so that, in 420 trials, there were only 5 failures to perceive heat. Small tested boys and girls from the 7th grade and the high school: in 21 trials, 5 reported heat, with no contact at all, 19 reported heat from the wooden point, while in 19 trials with the "heat-waves," 17 proved suggestible.

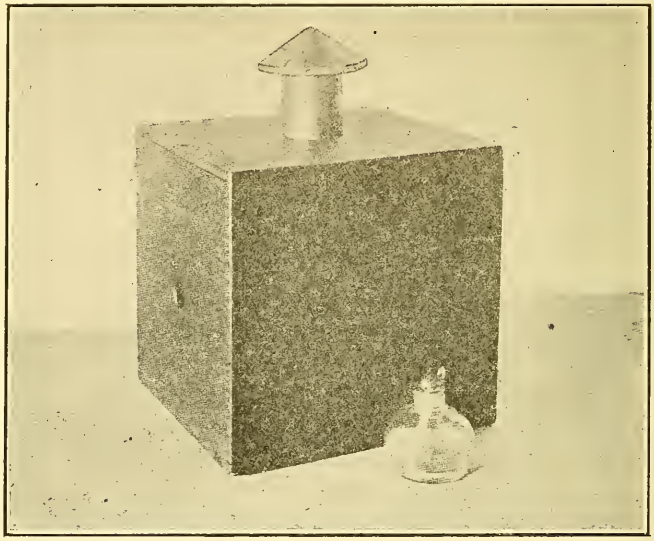

FIG. 76. GUIDI'S APPARATUS FOR THE WARMTH ILLUSION. (Modified by Whipple.)

C. ILLUSORY WARMTH—SMALL'S METHOD

Of Scott's 20 college students, 9 'yielded' 10 times (of a possible 10); 5 yielded 9 times; 2 yielded 4 times, and 1 each 8, 7, 5, and 3 times. No one of the 20 *S*'s resisted in every trial. Chojecki, who tested 30 men and 30 women students at the University of Geneva, got positive results from 19 (31.8 per cent.) with the use of Guidi's method. Okabe, who worked with school children and adults in the Cornell laboratory under the author's direction, obtained positive results in 70.7 per cent. of

the trials, and with 22 of 29 *S*'s (Table 126). The Italian children tested by Guidi were less suggestible (at least for his method), as Table 127 shows.

TABLE 126

Suggestibility to Warmth. Resistance-Coil Method (Okabe and Whipple)

GROUP	NUMBER	TRIALS	YIELDS	PERCENTAGE OF SUGGESTIBILITY	CASES WITH NO YIELDS
Men	12	59	43	73	1
Women	7	29	20	69	2
Bright boys	5	36	27	75	1
Dull boys	5	33	21	64	1
Totals	29	157	111	70.7	5

(2) The *relation to sex and to age* cannot be stated with assurance. Guidi's results indicate maximal suggestibility at the age of 9, but the Cornell tests, perhaps from being too few in number, failed to show characteristic differences between grammar-school boys and adults. It is likewise unsafe to generalize from the indications there given of the greater suggestibility of men.

(3) The *degree of suggestibility,* as indicated by the readiness with which warmth is felt, differs, as might be expected, in different *S*'s, *i. e.*, even of those who invariably perceive warmth, some report only "faint warmth," others "sudden heat," etc. Guidi classed his pupils into three groups, according as they took the suggestion quickly (in 1 to 2 sec.), moderately (2 to 3 sec.), or slowly (after 3 sec.), and found 33 per cent., 63.7 per cent., and 3.3 per cent. of his *S*'s in these three classes, respectively.

(4) Scott found no *correlation* between suggestibility as measured by the warmth illusion and suggestibility as measured by his flight-of-colors test, Chojecki no correlations between the results of his three methods, viz.: Guidi's 'stove,' Ochorowicz's 'hypnoscope' and Binet's progressive lines. Okabe's tests afforded the following low correlations with other forms of suggestibility tests: with progressive lines (Test 42)

0.17, with contradictory suggestion (Test 43) 0.21, with directive suggestion (Test 43) 0.29, with the weight illusions (Tests 40 and 41) none.

TABLE 127

Suggestibility to Warmth, as Related to Age. 187 Cases (Guidi)

Age	6	7	8	9	10	11	12	13	14	15
Per cent. suggestible	50	40.9	51.8	62.5	50	40	33.3	21.4	27.3	33.3

REFERENCES

(1) A. Chojecki, Contribution a l'étude de la suggestibilité. *ArPs(f)*, 11: 1911, 182-186.

(2) G. Guidi, Recherches expérimentales sur la suggestibilité. *ArPs (f)*, 8: 1908, 49-54.

(3) W. D. Scott, Personal differences in suggestibility. *PsR*, 17: 1910, 147-154.

(4) E. W. Scripture, Tests on school children. *EdR*, 5: 1893, 52-61.

(5) C. E. Seashore, Measurements of illusions and hallucinations in normal life. *SdYalePsLab*, 3: 1895, 1-67, especially 30-32.

(6) M. H. Small, The suggestibility of children. *PdSc*, 4: 1896, 176-220, especially 183-186.

CHAPTER XI

Tests of Imagination and Invention

Imagination, like most of the stock psychological terms, has the misfortune to be used in several different ways. In popular usage, imagination commonly implies something fanciful and unreal; we condemn a rumor, for example, by dubbing it "a mere figment of the imagination." In psychology, imagination has both a general and a specific meaning. Broadly speaking, imagination is equivalent to imaging, or thinking in images, as over against perceiving—re-presentation as contrasted with presentation. But the psychologist also differentiates between imaging which refers to some part of one's past experience (memory) and imaging, which, though necessarily based upon this same material, presents the material in new forms or patterns, and which is not felt to refer definitely to some part of one's past experience. This latter is imagination in the specific, or narrower meaning of the term.

A further distinction is made between imagination which occurs under passive attention, as illustrated in reverie, musing, or dreaming, and imagination which occurs under active attention, and which is marked by persistent, purposeful effort to dissociate former combinations of experience and to reorganize them into some new plan. We have, then, a distinction between passive imagination and active, creative, or productive imagination.

The tests of this chapter are designed both to secure indications of the wealth of spontaneous imagery in phantasy, and to measure capacity for creative or inventive thinking.

In so far as intelligence denotes not merely good attention and good memory, but also inventive capacity, ability to plan and organize, to anticipate, or to "put two and two together" (Ebbinghaus' *kombinierende Tätigkeit*), in so far must the attempt to measure intelligence employ tests of productive imagination and invention. It goes without saying that the tests here described do not exhaust the possibilities of investigation

in this important field of mental activity. Undoubtedly, new tests will be devised which will prove of value in supplementing those heretofore employed. We need especially a series of tests of inventive capacity, of graded difficulty, which shall put less emphasis upon linguistic attainments.

TEST 45

Ink-blots.—In their discussion of a proposed series of tests for the examination of individual differences in mental traits, Binet and Henri, in 1895, suggested that fertility of visual imagination might be investigated by means of a series of ink-blots. Two years later, but independently, G. Dearborn published brief suggestions for making a series of blots, and in the following year described the results of the use of 120 blots in the case of 16 Harvard students and professors. Since then Kirkpatrick has tried the ink-blot test with public school children of 8 elementary grades; Miss Sharp has followed the suggestion of Binet and Henri in a study of individual psychology upon graduate students in Cornell University, and Pyle has published preliminary averages for different ages.

The ink-blot test is commonly classed as a test of passive imagination, under the assumption that *S* simply looks at the blot and allows his associative processes to suggest to him whatever 'pictures' they may. In practise, however, *S* is quite likely to search actively for these associations, so that the mental activity concerned is, perhaps, more allied to active than to passive imagination.

Materials.—Standard series of ink-blots, numbered from 1 to 20. Stop-watch. Paper properly prepared for recording *S*'s statements.

The primary difficulty heretofore existing in the application of the ink-blot test has been the lack of standardized material. To meet this difficulty, the author has prepared the series of blots just mentioned by using zinc-block prints, so that investigators may now apply the same series of blots, and thus secure strictly comparable data. Unfortunately, this series has not yet been applied upon a sufficiently extensive scale to render it possible to publish norms of performance for the test.

Method.—(*a*) *Full procedure.* Instruct *S* as follows: "I have here a series of 20 odd-shaped ink-blots. I want you to

take them in order from 1 to 20, one at a time, to look them over at your leisure, and to tell me (or write down on a numbered blank) what things you can see in each blot. Try them in different positions. Of course, these blots are not really intended to be pictures of anything, but I want to see whether your imagination will suggest pictures of things in them, just as you sometimes try to see what objects you can make out of clouds." Let *S* take his own time. Especially with younger *S*'s, it is better for *E* to record the results, so that *S* may be perfectly free to enumerate as many things as are suggested to him.

Kirkpatrick used only four blots, and allowed each pupil one minute to name as many associations as possible for each blot. Miss Sharp used 10 blots, and allowed only 5 minutes for the (whole?) test.

The test may be conducted with a group of *S*'s by distributing the cards, and having them passed successively from member to member of the group until each *S* has written his associations for each card, but this method has obvious disadvantages.

(*b*) *Shorter procedure.* Following the method used by Dearborn, arrange the 20 cards face down in a pile, with the 20th card at the bottom, the 1st at the top, and the numbered edges toward *S*. Instruct *S* as follows: "Each of these 20 cards has on it an odd-shaped ink-blot. When I say 'now,' turn over the first card in this way [illustrating the movement that will expose the face of card No. 1 with the numbered edge toward *S*]. Look at the ink-blot, without turning the card in any other position, and say 'now' (or tap on the table) as soon as you have thought of something that the blot resembles. Of course, the blot is not really intended to be a picture of anything, but I want to see whether your imagination will suggest some 'picture' in it, just as you sometimes try to see what object you can make out of a cloud." Give the command 'now'; start the stop-watch at the same time. When *S* gives his signal, stop the watch, record the time and the object or association given by *S*. Continue in the same manner with the remaining cards.[1]

[1]The method proposed by Pyle (allowing 3 min. for writing the first thing suggested by each card in the order 1 to 20) is a modification of Dearborn's method for the purpose of making group tests. It suffers from the defects already pointed out (Ch. II, pp. 8-11) as characteristic of tests in which speed is made a measure of performance, in which written responses are introduced and in which a time-limit instead of a work-limit is employed.

Treatment of Data.—In the full procedure, the score is based upon the average or total number of associations; in the shorter procedure, upon the average speed of the single associations. It is also possible to form some estimate, in either case, of the type, richness and variety of *S*'s imagery by classifying the associations after some such plan as that illustrated below from Miss Sharp's results.

Typical Results.—The following associations for the 20 cards of the standard series are taken from the records of several adults, and will serve to indicate the variety that may be expected when the records of several *S*'s are compared. Note the frequent reference to animals.

(1) A lady seated on a couch. A witch riding on a new moon across the sky. A moose's head. A woman, sitting on a bank of shrubs, waving a handkerchief. Fir tree. Dragon in woods.

(2) Child, crouching in fear. Man with grotesque features. Ugly old colored woman. Old man seated. Back of bear. Lion crouching. Tree uprooted.

(3) A banner. A right-angled triangle. The God Billiken. An Egyptian idol. A jade-stone idol. A foot. Dog sitting on hind legs. Man shooting.

(4) A large beetle. A boat load of excursionists. A lobster. A spider. Potatoes. A dirigible balloon of the Zeppelin type, with a cloud of steam or smoke overhead, and a grappling-anchor trailing below. Two trees and roots. Stockings on clothes-line.

(5) A pig. A woman with a big head of hair. A butterfly. A hole through the ice. A girl wearing a tam-o-shanter cap. Human liver and heart. A rock. An oyster shell.

(6) Woman running and holding her skirt. Woman with a muff in her left hand, and her hat almost blown off. A broken bellows. Merry Widow waltz. A dog on a post. An island and lake.

(7) Large caterpillar on a horse's shoulders. A devil bending over something. An old man. A dream monster. A woman with flowers. Unicorn. Pig.

(8) Human torso. Hot and cold water faucet in a bath tub. Person with head bent forward, holding sticks in her hand.

Heads of two birds trying to swallow what is between them. A frog. A vase. Vertebra of back bone.

(9) A goat with a pack on his back. A turkey with drooping wings trailing on the ground. A tree. A goose's head.

(10) Ugly man's head. Head and arm of a woman with a lighted candle in her hand. A dachshund running off with some one's cape. A mosquito pupa. A tree.

(11) Map of Scotland and Ireland. Owl that has just placed a fish before him on the branch of a tree. Some specimen in geology. A tree blown in a heavy gale. A tiger under a tree. A conch shell.

(12) Map of United States and part of Canada. A chicken lying on its back. An Indian head. A woman sitting on a cliff under a tree, reading a novel. A buffalo running. Hydroids.

(13) A flying squirrel. The skin of a bear. A hen sitting on a nest. A dog running.

(14) A crab. A bat with outspread wings. A moth. A neurological slide. A flower. An insulator.

(15) Section of medulla oblongata. Two nuns bowing their heads together. A tulip. A false mask. A crab. Head of a fish. A plate of false teeth. A design-unit of two bears with heads together.

(16) Closed hand with thumb and little finger, or a sixth finger, projecting. A loving cup. A tea-pot. A head.

(17) A root. A porcupine. An Indian head. A nerve cell. Sponge dropping water.

(18) A Chinese dragon, as seen on packages of fire-crackers. Branch of a gnarled oak. A lizard. An old woman and child. A man with knees bent.

(19) Bird alighting on a nest. A flying squirrel. Rear of a cat in rapid motion. A lamb. A duck.

(20) Man pulling off his sweater. Runner leaning forward to start a foot-race. Photographer, with focussing cloth over his head. Crocodile suspended by the head. Bear with the grandmother's night-cap and gown, as illustrated in Little Red Riding Hood. An elephant seated.

General Results.—(1) *Speed of association.* In 1920 trials, Dearborn found the average time for making a single association to a blot to be 10.3 sec. This seemingly long time may be due to the difficult nature of some of the blots in his series.

The children aged 8 to 14 tested by Pyle with the author's blots, but with the written response, averaged from 6.4 to 12.0 responses in 3 min. His adults averaged 10.6 for the men and 9.8 for the women.

(2) *Dependence on age.* Kirkpatrick states that "younger children seemed more suggestible or imaginative, as they named more spots" (Table 128). Pyle's tables show a similar tendency.

TABLE 128

Average Number of 'Names' Given to Ink-Blots (Kirkpatrick)

GRADE	I	II	III	IV	V	VI	VII	VIII
Average	2.9	2.5	2.6	1.8	1.9	1.7	2.1	2.2

It is evident that something besides a simple decline of 'imagination' with age is exhibited in this table. In explanation, Kirkpatrick says: "The younger children seemed to have no doubt whatever of the spot being a picture of the object they named, while the older children simply said 'it is some like' or 'it looks a little like,' 'a dog,' 'cloud,' or whatever else was suggested. This superiority of the small children is striking when we consider that the number of mental images that they have is much smaller than that possessed by older children, who may name a part of the body or the map of a country or something else that the younger children know nothing about.

"The smaller number of objects seen in the spots by the children of the 4th, 5th, and 6th grades is probably to be explained by the fact that children of those ages have become more critical in their sense-perception, as their ideas have become more definite, and as they have learned from life's experiences and from training to be more careful in their judgments. The older pupils of the 7th and 8th grades, on the other hand, have passed into another stage in which they realize that a picture is not necessarily this or that, but may resemble any one of several things, hence they are not afraid to say what it looks like."

(3) *Dependence on occupation.* Dearborn believes that, at least in maturer *S*'s, the results of the ink-blot test are conditioned, not so much by age or sex directly, as by habits of living, occupation, and other environmental factors: thus, we should

expect characteristically different results from the test when applied, for example, to artists, farmers, laborers, professional men, to the city-bred or the country-bred, etc.

(4) *Dependence on race.* Pyle's averages (5b) show that negroes are nearly as good as whites in the ink-blot test.

(5) *Individual differences*, both in speed, number and type of association seem to have been been clearly marked and fairly constant, whenever the test has been applied. Thus, in Dearborn's single-association method, the highest agreement in the answers of his *S*'s for any one card was but 40 per cent., while for several cards, no two *S*'s gave the same answer.

As regards fertility of imagination, Miss Sharp noted that the most imaginative *S* in her group saw 81 objects, the least imaginative but 27 objects in the same 10 blots. The same investigator believes, however, that all *S*'s might be roughly divided into two groups, (*a*) the constructive or imaginative, who put together concrete details "in such a way as to form a significant whole," and (*b*) the matter-of-fact, or scientific type, given more to analysis than to creative synthesis.[2]

As examples of this difference, the following reports from two of Miss Sharp's *S*'s may be quoted: both refer to the same blot.

(1) *Associations few and non-constructive.* "An eagle. Stuffed turkey. Head and neck of a musk-rat."

(2) *Associations numerous and constructive.* "Giraffe. Prehistoric bird in flight. Fairy riding on a bumble-bee. Bit of tropical jungle, with trailing gray mosses and pools of water. Japanese lady. Bit of landscape with two hills and a valley between—an army encamped under one hill. Moss-grown log floating in water. Fabulous monster (griffin perhaps) walking off on his hind legs with a small Hottentot under his arm."

(6) *Qualitative classification.* It is often possible to classify the associations peculiar to a given *S*. Thus, Miss Sharp mentions as classificatory groups: (*a*) common-place, every-day objects, such as domestic utensils, tools, plants, and particularly animals, (*b*) scientific objects, such as geometric figures, schematic drawings, (*c*) objects suggested by literary reminiscence, and (*d*) objects from fable and mythology, such as

[2]It is tempting to regard this classification as identical with the common classification of laboratory *S*'s into 'subjective' and 'objective' observers.

centaurs, dragons, witches, fairies, etc. Some *S*'s exhibit variety of association, in that they cite objects that belong to several of these groups; others are much less fertile in imagination and confine themselves largely to a single type of imagery.

REFERENCES

(1) A. Binet et V. Henri, La psychologie individuelle. *AnPs*, 2: 1895 (1896), 411-465, especially 444.

(2) G. Dearborn, Blots of ink in experimental psychology. *PsR*, 4: 1897, 390-1.

(3) G. Dearborn, A study of imaginations. *AmJPs*, 9: 1898, 183-190.

(4) E. Kirkpatrick, Individual tests of school children. *PsR*, 7: 1900, 274-280.

(5) W. H. Pyle, (*a*) The examination of school children. New York, 1913, pp. 33-35. (*b*) The mind of the negro child. *School and Society*, 1: 1915, 357-360.

(6) Stella E. Sharp, Individual psychology: a study in psychological method. *AmJPs*, 10: April, 1899, 329-391.

TEST 46

Linguistic invention.—The ink-blot test serves primarily as a test of visual imagery. But an even more fruitful source of individual differences in creative ability may be found in linguistic invention. Miss Sharp, acting upon the suggestions of Binet and Henri, tested what she terms 'literary imagination,' in three ways, viz.: by the development of sentences, by the development of a given theme, and by the choice of a topic for composition.

The idea of presenting a number of words to be joined into a sentence has been elaborated in various ways. The assignment of three words was employed by Masselon in 1902, and this test has, on that account, been referred to by some writers as the "Masselon method." It forms one test in the well-known Binet-Simon series, and was one of the tests used by Miss Sharp in her investigation of the mental types of adult *S*'s. The reduction of the number of terms supplied, to two has been strongly recommended by Meumann, who selected the two terms in a special manner (see below), while another variation of the two-word test has been tried by Burt and by Wyatt, in which *S* is given a series of 10 words to be joined together successively, by pairs, in a series of sentences. On the other hand, the number of terms has been increased to 5, 8 or 10 with the in-

struction to invent a story containing the prescribed words (invention of stories). This method evidently stands midway between the method of sentence-formation and the method of development of a theme, while by a little further extension the well-known Ebbinghaus completion method (Test 48) is reached. It needs little reflection to understand that the nature of these various tests becomes decidedly varied as the number and nature of the supplied terms is varied.

The method of completing a prose passage in which a large amount of the original text is supplied is embodied in Test 48. The present test includes the method of sentence-formation known as Masselon's method, the method of sentence-formation devised by Meumann, the completion of sentences used by Binet, the invention of stories, and the development of a theme.

Before undertaking these formal tests, however, it is desirable, if the purpose in mind is to make a qualitative study of the mental type of individual *S*'s, to institute a preliminary inquiry concerning the general literary tastes and habits of each *S*. The exact nature of this inquiry must, naturally, be adapted to the age and training of the *S*'s: the following are some of the points that have been covered by investigators: (1) list of favorite books, (2) statement of favorite type of reading, (3) statement of the magazines, periodicals, newspapers, etc., ordinarily read, (4) list of books (outside of classroom or professional work) read during the last year, (5) statement of favorite games and evidence of enjoyment of games, like chess and checkers, that demand creative activity and foresight, (6) fondness for the theater, drama, music, painting and other forms of art, etc., (7) experience in creative literary work.

A. SENTENCE-FORMATION (MASSELON METHOD)

METHOD.—Ask *S* to write as many sentences as possible containing the three nouns: *citizen, horse, decree.* Each sentence must contain all three nouns, though it may contain others as well. The sentences are to be as varied as possible. Five minutes are allowed. Continue the test with four more sets of nouns, and afterward make similar tests with five sets of verbs.

For the noun tests, use as additional sets: (2) *bell, ground, owner,* (3) *skill, modification, picture,* (4) *cup, fraction, money,* (5) *letter, law, summer.* For verbs use (1) *bless, destroy, write,* (2) *make, correspond, remain,* (3) *require, choose, run,* (4) *see, find, throw,* (5) *remember, put, depart.* In the noun tests, *S* is permitted to use either singular or plural forms, and possessive as well as nominative or objective cases; in the verb tests, he may use any form of the given verb, *e. g., blessed, to bless, will bless,* etc., as well as *bless.*

The tests may be conducted with individuals or with groups; but it is preferable, especially with young *S*'s, to work individually and to let *S* dictate the sentences instead of writing them.

Treatment of Data.—The quantitative score is determined by averaging the number of sentences written by *S*. The quality of work may be graded upon any convenient scale, *e. g.,* 1 to 5, corresponding to five degrees of excellence. Miss Sharp used the symbols *A, B,* and *C,* and indicated intermediate grades by the use of — and +. For purposes of computation, she then assigned numerical values to these symbols, as follows: *A*— = 40, *A* = 50, *A* + = 60, *B*— = 80, *B* = 100, *B*+ = 120, *C*— = 160, *C* = 200, *C* + = 240. In practise, this scoring is virtually equivalent to estimating quality of work in terms of average number of words per sentence, and that simpler method may be used for the qualitative score.

Typical Results.—(1) The following are selected single sentences reported by Miss Sharp for the first test:

1. "Decrees are made for citizens, not for horses." (The connection of the words here is simple and mechanical.)
2. "That stalwart citizen on the great gray horse is a man to be trusted with the decree." (This implies a concrete situation.)
3. "All the well-to-do citizens of the village, each mounted on a horse, rode through the streets, proclaiming their dissatisfaction with the new decree." (A situation is here more fully outlined.)

(2) The following is a full set of sentences written by a graduate student, in 5 min., for the first assignment:

1. A decree was posted that the citizen should not abuse the horse.
2. The horse of the citizen was sold by official decree.
3. "Here," said the citizen, "is the horse mentioned in the decree."

4. Early in Arabian history, a decree raised to a higher caste, a citizen who owned a horse, but later, possession was sufficient for better standing, and the law was not needed.
5. If a citizen keep a horse, it is a decree that he use it kindly.
6. "What a funny decree," exclaimed the citizen, when he read of the horse sun-bonnet law.
7. The decree was signed that the horse had kicked the citizen, and therefore the injured man could collect damages from the owner of the animal.
8. "Time is up," cried the citizen, stop-watch in hand, "I hereby decree that you write the word horse and stop at once." [Faulty on account of the use of 'decree' as a verb.]

(3) The following represent groups of sentences written for the author by two college students (selected at random from a number of papers) for the fourth set of verbs. The relatively greater variety of the second group is clear.

A. 1. "I saw the book and tried to find a place in which to throw it."
2. "I threw the cat in the creek and turned to see if anyone had found me out."
3. "I see that I can find nothing to throw at him."
4. "You see, it was this way, I simply found the hatchet and threw it."

B. 1. "The child saw a horse, found a stone and threw it at him."
2. "When you find a clover, see if it has four leaves: if not, throw it away."
3. "Throw the paper out of the window and see if it will find a good landing place."
4. "Find me a pencil, then I will see if I can find out the solution to the problem which is on the paper that you threw into the basket."
5. "The boy found an apple, but when he saw it was decayed, he threw it away."

CONCLUSIONS.[1]—(1) *Dependence on part of speech assigned.* All *S*'s tend to write fewer, but better sentences with verbs than with nouns.

TABLE 129

Scores of Seven Adults in Developing Sentences (Sharp)

FORM OF TEST	QUANTITY OF WORK			QUALITY OF WORK		
	Average	Maximum	Minimum	Average	Maximum	Minimum
Nouns 'given'---	4.6	6.6	3.2	79	113	55
Verbs 'given'---	3.8	5.8	2.5	93	133	54

[1]These are all drawn from the work of Miss Sharp.

(2) The rank of *S*'s, both in quality and quantity of work, is the same when nouns and when verbs are assigned.

(3) "In general, the subjects who made the most sentences made the most elaborate, and those who made the fewest sentences made also the simplest and most unimaginative."

(4) This test correlates with the ink-blot test, in so far as those *S*'s who show most constructive capacity with the blots also show most constructive capacity in the development of sentences.

B. SENTENCE-FORMATION (MEUMANN'S METHOD)

The Masselon method, according to Meumann, is less well fitted to bring out differences in intelligence than his own method of presenting but two words, so selected that a number of different relations can be worked out between them, only one of which, or at least only a few of which, can be regarded as being really appropriate, pertinent and sufficiently definite as to evince good sense and a real appreciation of the relation. This appropriate combination of the two words into a sentence is accomplished only when *S* introduces a third relational element that supplies the 'point' needed to round out the thought.

METHOD.—Explain to *S* that he is to make a sentence with each pair of words. By the aid of illustrative examples make it clear that there are two ways in which any pair could be joined, the one correct enough, perhaps, but banal and loose, the other logical, sensible and specific, and that the latter form is the one desired. For example, the words *snow—melts* could be rendered as "The snow melts" or as "Snow melts when the warm sun shines on it." Again, the words *square—sides* could be rendered as "A square has sides" or as "A square has four sides of equal length." Similarly, from the pair *automobiles—tires* could be obtained "Automobiles have tires" or "Automobiles have pneumatic tires to make them ride easily."

When these instructions have been grasped, give *S* the following 10 pairs of terms and allow him all the time he desires to write one sentence of the 'pointed' type for each pair: (1) donkey—beatings, (2) soldiers—country, (3) city—streets, (4) sun—noon, (5) pine—winter, (6) drink—poverty, (7)

cat—punished, (8) sky—red, (9) water—hill, (10) teacher—pleased.[2]

RESULTS.—The following types of answers may be readily distinguished:[3]

(*a*) The dictated words are written, but nothing else.

(*b*) The given words are joined in a nonsense statement, *e. g.*, "The city is a street."

(*c*) A number of successive sentences are cast in a very simple form which is the same in each, *e. g.*, "A donkey has beatings." "Soldiers have a country." "A city has streets."

(*d*) The written statement is incorrect, but such as to suggest that *S* had the glimmerings of an idea that failed of expression, possibly on account of some difficulty in the use of language, *e. g.*, "To drink is poverty."

(*e*) The sentence is logically correct, but indefinite, too loose, general and banal, *e. g.*, "A city has streets."

(*f*) The words are combined into a specific statement, but one that is imaginative and not expressive of the correct connection, *e. g.*, "Once upon a time there were three soldiers who lived in a beautiful country."

(*g*) The sentence is definite, logical, correct and pertinent, embodying the right causal connection, *e. g.*, "In the city the streets are wide and paved with brick." "Good soldiers are ready to die for their country."

NOTES.—The 'sentence-construction' or 'sentence-formation' test used by Wyatt and by Burt consists in presenting a series of 10 words such as *circle, moon, night, sleep*, etc., each one of which is fairly obviously connected with the next and then allowing each *S* 2.5 min. to write a series of sentences connecting the successive terms by pairs, *e. g.*, "The full moon has the form of a circle." "The moon shines at night," etc. Particular stress is laid upon the condition that the various sentences

[2]The last two have been supplied by the author to replace less useful or more complex combinations in Meumann's list.

[3]Although Meumann concludes that any attempt to score this test quantitatively must be arbitrary, it would seem possible to attempt some numerical comparison of the work of different *S*'s by assigning a scale of marks, like 0, 1, 2, etc., for these several qualitative degrees of performance.

must show the "closest possible connection." In practise this instruction is difficult to make clear, and the performance of *S*'s is quite difficult to score precisely or fairly. The time consumed in writing also enters as a disturbing factor. Burt found for this test a coefficient of reliability of only .61, but a fairly high correlation with intelligence, 0.62.

c. COMPLETION OF SENTENCES

MATERIALS.—Printed forms containing beginnings of 25 sentences,[4] with spaces for the completion of each sentence. Piece of white cardboard. Stop-watch.

METHOD.—Give *S* the following instructions: "On this paper there are printed the beginnings of a number of sentences. I am going to show these to you, one at a time. As soon as I show you one, I want you to finish out the sentence. You may say anything you want to, as long as the whole sentence will make sense when you have finished it. Take an easy attitude toward the test. Don't try to hurry. Let the completion of the sentence develop naturally and freely, whether it is long or short." If *S* fails to understand what is wanted, supply him with an extra paper on which a few trial sentences have been written in pen and ink, and show him how they might be completed. For the test proper, cover the entire test-blank with the cardboard: after a warning 'ready,' expose the first incomplete sentence.[5] Start the watch at the same time. Record as nearly as possible the time used by *S* in starting to complete the sentence, *i. e.*, the time he takes, after he reads the sentence, to 'get an idea.' The timing should be done without *S*'s knowledge.

VARIATION OF METHOD.—The printed forms are arranged to permit written tests, either of individuals or of groups. With groups the timing may be omitted without serious detriment.

[4]The first 20 sentences are taken, with such slight modifications as translation has suggested, from Binet. The last five (since Binet prints but 20 of the 25 he recommends) have been supplied by the author. Other sets of incomplete sentences will be found in Weidensall or in Woolley and Fischer.

[5]The sentences have purposely been numbered from the bottom of the page, so that the cardboard will not interfere with *S*'s writing.

Some *S*'s give shorter, others longer sentences when they are written.

TREATMENT OF DATA.—Compute the average, or determine the distribution, of the times needed by *S* to start the 25 sentences. For a qualitative index, estimate as well as possible (preferably by using some such system of scoring as that described in the development-of-sentences test) the general value of the completed sentences. *S*'s sentences may also, if desired, be classified in regard to type, *e. g.*, vague or meaningless, commonplace, reminiscential, imaginative, aphoristic, etc.

A more elaborate system of scoring was attempted by Woolley and Fischer and followed by Weidensall in part. Records were kept of (1) number of sentences attempted, (2) number of sentences correct (in the sense of constituting a real sentence, even though there might be some mistakes of grammar), (3) number of simple and of complex sentences, (4) average number of words written per sentence, (5) number of ideas expressed in the sentences, taken collectively (scored by a somewhat complex set of rules), (6) total time used in the test, (7) time used to start each sentence (classed in five groups, 0-2, 3-5, 6-10, 11-20 and 21-60 sec.), and (8) 'index' of ideas, obtained by dividing (6) by (5). Use is made in the published results, however, of only the 2d, 5th, 7th and 8th of these scores.

RESULTS.—(1) Binet found characteristic differences in the *speed of work* of his two daughters, Armande and Marguerite. Thus Armande's records show 12 sentences started in less than 5 sec., 4 sentences in from 5 to 10 sec., 6 in from 10 to 20 sec., 1 in 28 sec., and 1 in 70 sec. Marguerite's records, on the other hand, show but 1 sentence started in less than 5 sec., but 7 sentences in less than 10 sec., and the remainder in much longer times, *e. g.*, 20, 50, and 70 sec.

(2) Binet's two *S*'s also showed characteristic differences in the *type of sentence-completion:* Armande is poetic and imaginative; Marguerite's sentences are more precise, more practical, more in accord with real life, less emotional. For example, for Sentence 1, Armande writes: "I entered the field by a covered footpath." Marguerite writes: "I entered the grocery and bought two cents worth of chocolate."

(3) The study of school children at Cincinnati by Woolley and Fischer brings out the following points: (*a*) *age* is a factor of some moment, since 15-year-old pupils, when compared with 14-year-old, show a decided improvement in number of correct sentences, a large increase in number of ideas expressed and a slight increase in the speed of beginning the sentences; (*b*) *sex differences* seem to favor the boys, who are somewhat superior to girls in correctness and somewhat quicker than girls in speed of beginning (there was no definite sex difference in number of ideas); (*c*) the test shows a large positive *correlation with school grade* attained by both sexes at both years, when performance is scored by any of the three measures—number of correct sentences, number of ideas or speed of response.

(4) *Delinquents.* The results obtained by Weidensall with Bedford Reformatory women show that they are slower to respond than the Cincinnati girls, slower even than the Cincinnati retarded girls. On the other hand, the number of correct sentences and the number of ideas expressed were, contrary to expectation, greater in the Bedford group; this outcome may be due to the conditions under which the Bedford women were tested, or it may be connected, one may surmise, with the longer time taken in starting the sentences. At Bedford the ability to make correct sentences did not correlate with school grade attained before entering the institution, but the speed with which the sentences were started did correlate with the school grade; in fact, the poorest *S*'s took five times as long to start their sentences as did the more intelligent ones.

D. INVENTION OF STORIES

Test No. 26 of the Binet-Simon 1905 series called for the construction of a sentence containing three specified words. This test has been elaborated by Mrs. Squire by asking not for a sentence, but for a *story* about three words. The same test, with 5, 8 or 10 words given, has been used by Winch in his comparative study of memory for ideas and productive imagination.

Meumann's somewhat similar test consists in dictating a series of 'cue-words' or phrases, carefully selected as to nature

and number, with the instructions to make a story from them. In this test the principles involved in selecting the words make the problem of a different sort from that involved in Winch's test; the conditions are rather more rigorously drawn, so that the number of satisfactory solutions is smaller. In some respects, in fact, Meumann's test more nearly resembles the Ebbinghaus completion method (Test 48).

METHOD.—(1) For young *S*'s, ask for a story about a *boy,* a *river* and a *ball* (Squire test). For children younger than 10, and better for yet older children, the story should be given orally and taken down verbatim by *E.*

(2) Winch's instructions, as given in writing to a group of 13-year-old *S*'s, were as follows (8, p. 102):

"Write a story containing the following words: *thief, landlord, crab, shake, hotel, basket, cries, provisions, escape, custody.*

"You are to write the longest story you can, because the longer the story is, the more marks you will get, provided that everything you write has something to do with the story. You will get no marks at all for them and only be wasting your time if you write sentences which have no connection with the rest. Try and think out the story you are going to write before you start, and see that the progress of the story will enable you to fit all the words in properly."

It is desirable to make more than one test of this sort. For this purpose, use may be made of one or more of the other lists of terms used by Winch[6] with the same instructions as above. These lists are: (1) *Orphan, garden, hungry, station, parents, clothing, visitor, cottage, train, country.* (2) *Snowstorm, children, ticket, clock, dog, screams, church, basket, river, ice.* (3) *Army, hill, artillery, victory, cavalry, fight, captured, brave.* (4) For younger children (8-9 years): *dog, clock, basket, man, children.*

(3) For Meumann's test *E* must take a simple connected bit

[6]It should be said that Winch's invention tests were applied to children who were also being tested in 'substance memory' with the aid of passages containing terms quite similar to those given as material for the invention. This had undoubtedly an effect upon the invented stories, though Winch asserts that "the invented stories are, almost invariably, on a much lower plane" (p. 101).

of prose depicting a total situation and reduce it to a series of salient cue-words. After giving one or more preliminary illustrations, *S*'s are requested to write a story based on the cue-words selected. The following is the set most successfully used by Meumann: *house took fire—child alone—clever monkey—parents thankful—reward.*

No time limit is set in any of these tests. *S*'s should not be hurried.

Treatment of Data.—Mrs. Squire contented herself with recording four degrees of performance in her three-word test: (*a*) complete failure, (*b*) separate sentence given for each word, (*c*) three words in one sentence, but the sentences [of the rest of the story?] unconnected, and (*d*) complete narrative. The outcome of this scoring is indicated below.

Winch scored performance on the general basis of number of meaningful 'units' in the story, giving no allowance for any sentences or parts of sentences which did not arise connectedly from preceding sentences, but yet no penalizing for lack of esthetic unity (making every element in the story converge to a point). The aim is to rank the performance with regard to the "fertility of continuous and connected imagination" displayed in it.[7]

Results.—(1) *Dependence on age* is shown in Mrs. Squires' results in the form of (*a*) "a development from the crude sentence strung together by 'ands' to a closely knit sentence," while (*b*) "another characteristic change is the transition from the fantastic type of story related by the 6, 7 and 8-year-old children to the extremely realistic, matter-of-fact style employed by the 9th, 10th and 11th year groups," and (*c*) "another plan of invention, more flexible in style is evident in the stories of the 12th and 13th year groups."

As applied in her mental age series, this test becomes roughly diagnostic as follows: the normal 6-year-old can give orally sentences containing the three words; ability to get all three words into one sentence, though with a disconnected story, would appear typical of 8 and 9-year-old children (though given by Binet as a 10-year test); ability to construct a complete

[7]See his discussion, pp. 102-105, for further details.

narrative as a realistic type is seen in the 9th and succeeding years, with a final level of flexibility and superiority of style appearing at 12 and 13 years.

(2) *Correlations.* Winch found moderately high correlations, .55 to .75, between substance memory and the invention of stories, and that practise in substance memory, unless carried to the 'fatigue-point' (limit of training), tends to augment proficiency in invention.

(3) *Reliability.* The reliability of the test, as might be anticipated, is not very high, about .50, so that more than one trial is demanded for significant results.

(4) *Qualitative differences.* Meumann found it somewhat difficult to score the work of school children in such a way as to distinguish fine degrees of intellectual ability, but he considers the elaboration of the story from the cue-words a good test for revealing larger differences in general mental ability and also for revealing various mental types. In general, he finds eight fairly distinct types of story.

(*a*) Zero performance; connections between the cue-words lacking or nonsensical.

(*b*) The cue-words are connected in a number of separate and independent sentences. Here the grade of mental ability is sufficient to join together pairs of terms, but not to make the larger synthesis of all the terms into a whole.

(*c*) Attempts are made to produce a whole, but the connections between the various cue-words are not rightly arranged and the point of the whole story is not grasped.

(*d*) The connections between the cue-words are rightly arranged, but the point of whole series of words is missed and the result is a story of a totally wrong 'turn.'

Types *a-d* may be regarded as all indicative of lower stages of intellectual ability: the following four types, however, may be regarded as solutions of the problem, though of different kinds.

(*e*) The *pure imaginative type* is illustrated by a story of astonishing richness of detail, with decided linguistic fluency, but with the real point either quite lost or badly distorted. *S*'s of this type evince, then, little intelligence, but a rich imagination; their endowment is perhaps exclusively linguistic.

(*f*) The *pure intellectual type* is illustrated by a story in which the connections of the cue-words are correct and the point of the whole is correctly grasped, but its elaboration into a story is accomplished in the scantiest manner possible. *S* is content to present the logical and factually correct connection of the cue-words in the shortest possible manner.

(*g*) The *imaginative-emotional type* is illustrated in stories that show evident presence of feeling, to give due expression to which *S* indulges in active imagination. He introduces invented details to express his emotional reaction.

(*h*) The *intellectual-imaginative type* is illustrated by stories that

show clear grasp of the meaning of the whole supplemented by imaginative and pictorial additions, which are, however, always pertinent and subdued to the salient points in the development of the story.

For examples of these several types, consult Meumann (pp. 158-9): a single one of them, that of Type f, may be repeated here:

Story of Arthur W., 7th school year, age 12 years 9 months:

"There was a house in the village: it took fire. The parents had just gone away. The child was all alone in the house. The people also had a clever monkey. He saved the child from the flames. And when the parents came home and saw that the monkey had saved the child, they were thankful and gave him a reward."

E. DEVELOPMENT OF A THEME

METHOD.—Supply *S* with writing materials, and give him 10 min. (or perhaps longer if working with young *S*'s) to write upon some theme selected from the following: (1) *The Death of a Dog,* (2) *The Capture of a Fortress,* (3) *The Escape of a Prisoner,* (4) *A Forest Fire,* (5) *The Mission of Music,* (6) *The Influence of Newspapers,* (7) *The Delays of Justice,* (8) *A Trip in a Flying Machine.*[8]

TREATMENT OF DATA.—Quantity or speed of work may be reckoned with approximate accuracy by counting the number of words written in the assigned time; quality of work, which is really important, especially in the treatment of imaginative themes, must be estimated by *E* after a trial has shown what may be deemed poor, and what good work for *S*'s of the age under investigation. Quality may be recorded in the manner already described, or upon the basis of 100, as in grading school compositions.

RESULTS.—(1) The relative number of ideas elaborated by different *S*'s is indicated with fair approximation by the relative number of words written, so that number of words may stand as a fair index of fluency of ideation and general linguistic readiness.

---

[8]The first of these themes was used by Binet in his comparative study of the mental processes of his two daughters; the next six were used for a similar purpose in Miss Sharp's study of university students—the first three of them being designed to involve imaginative, the second three expository treatment; the last theme is suggested by the author as more suitable for younger *S*'s. To secure a more reliable estimate of *S*'s efficiency it is desirable that more than one theme should be developed.

This test lends itself readily to group treatment, since it involves a familiar type of school activity.

(2) As a rule, more words are written upon imaginative than upon expository themes. Sharp's best *S* wrote in 10 min., on an average, 259 words upon imaginative, and 222 upon expository themes; her poorest *S* wrote, on the average, 124 and 94 words, respectively, for the same types of themes.

(3) Those *S*'s that show constructive ability in the ink-blot test, and in the development of sentences, also exhibit the same superiority here in the development of themes.

NOTES.—These tests of linguistic invention might, without great difficulty, be paralleled in other fields of constructive effort. A test of musical ability (of the creative sort) might, for example, be devised by asking *S*'s to finish a partially given musical theme, or to construct a simple melody from a given series of notes. Similarly, certain forms of artistic invention might be tested by asking *S*'s to sketch designs for wall-paper or patterns for Venetian iron-work.

Miss Sharp's test of the choice of a theme was conducted by asking *S*'s to select, from the following 10 themes, those five upon which they would prefer to write, if asked to do so: A. Imaginative themes, (1) *In a Snowstorm,* (2) *A Polar Landscape,* (3) *A Puritan Sabbath,* (4) *My Opposite Neighbor,* (5) *Man Endowed with the Power of Flight:* B. Expository themes, (6) *Civilization not Regeneration,* (7) *Wisdom in Charity,* (8) *Friendship of Books,* (9) *Fiction as a Vehicle of Truth,* (10) *The Eloquence of the Bar and that of the Pulpit.* The expository themes were generally preferred, but some *S*'s, who, as other tests showed, had little capacity to handle imaginative themes, did select several from the latter division.

REFERENCES

(1) A. Binet et V. Henri, La psychologie individuelle. *AnPs,* 2: 1895 (1896), 411-465, especially 444.

(2) A. Binet, L'étude expérimentale de l'intelligence. Paris, 1902. Pp. 309. Especially ch. x.

(3) C. Burt, Experimental tests of higher mental processes and their relation to general intelligence. *JEPd,* 1: 1911, 93-112.

(4) R. Masselon, Psychologie des déments précoces. Paris, 1902.

(5) E. Meumann, Ueber eine neue Methode der Intelligenzprüfung und über den Wert der Kombinationsmethoden. *ZPdPs,* 13: 1912, 145-163. (Also Vorlesungen zur Einführungen in die experimentelle Pädagogik, 2d ed., Vol. 2: 1913, 445-452.)

(6) Stella E. Sharp, Individual psychology: a study in psychological method. *AmJPs*, 10: 1899, 329-391.

(7) Carrie R. Squire, Graded mental tests. *JEdPs*, 3: 1912, 363-380, 430-443, 493-506, especially 439-441.

(7a) Jean Weidensall, The mentality of the criminal woman. To appear in *EdPsMon*.

(8) W. H. Winch, Some relations between substance memory and productive imagination in school children. *BrJPs*, 4: 1911, 95-125.

(8a) Helen T. Woolley and Charlotte R. Fischer, Mental and physical measurements of working children. *PsMon*, 18: 1914, No. 77. Pp. 247, especially 185-212.

(9) S. Wyatt, The quantitative investigation of higher mental processes. *BrJPs*, 6: 1913, 109-133.

TEST 47

Word-building.—The word-building test was suggested by the familiar game of anagrams, as well as by the advertisements often seen in magazines in which a prize is offered to the person who can make the most words from a given word or series of letters. This test is easily administered and evaluated; it is one that calls for ingenuity and active attention; it might fairly be said to demand that ability to combine isolated fragments into a whole, which Ebbinghaus has declared to be the essence of intelligence and for the measurement of which he devised his well-known 'completion method' (Test 48); and finally, its execution is conditioned to a certain extent by the richness and readiness of the examinee's word-vocabulary. One may expect, therefore, to find a correlation between this test and the vocabulary test (No. 50), and possibly between it and school standing or general intelligence, and other tests of creative literary ability.

In addition to the preliminary reports made by the author, the two tests proposed by him have been tried out by Pyle, Squire, Wyatt (with some modification) and most recently by Anderson, in an extended application to several hundred public school children at Ithaca carried out under the author's direction for the purpose of supplying curves of percentile distribution for performance in several tests. Heymans and Brugmans have used a similar test (making as many words as possible in 10 min. from a given 10-letter word) in their study of the intercorrelations of various tests of intelligence.

MATERIALS.—Two specially prepared blanks, the first of which calls for combination of words from the letters *a, e, o, b, m, t*, the second from the letters *e, a, i, r, l, p*.

METHOD.—Provide *S* with the first test blank, and give him the following instructions: "Make as many words as you can from the six letters given on this blank. You may use any number of letters from one to six, but no letter may be used twice in the same word, and no other letters than these six are to be used. You will have five minutes." Conclude the test by use of the second blank under the same conditions. For comparison with the curves of distribution given here, both tests must be applied and in the order just mentioned.

TREATMENT OF DATA.—Each word written in accordance with the rules counts one. To determine just what shall be termed a 'word,' the data secured by Anderson, upon which the curves that follow are based, was scored by following the division made upon each page of the 1910 edition of Webster's New International Dictionary, *i. e.*, any word found above the line was admitted; any word found below the line (and hence rare, obsolete, dialectic, etc.) was excluded. No discount was attempted for possible instances in which legitimate words were hit on by mere accident. As a guide to scoring these tests, the lists of admitted words are reproduced here.

aeobmt-test

a	bam	ea	ma	o	tab
ab	bat	eat	Mab	oat	tambo
Abe	bate	eta	Mae	om	tame
abet	be		mao		tea
am	beam		mat		team
ambo	beat		mate		to
at	bema		me		toe
ate	bet		meat		Tom
atom	beta		met		tomb
	bo		meta		tome
	boa		moa		
	boat		Moab		
	bot		moat		
	bote		mob		
			mot		
			mote		

cairlp-test

a	ea	I	la	pa	Ra
ai	ear	Ira	lair	pail	Rae
ail	earl	ire	lap	pair	rail
air	ela		lea	pal	rale
al	Eli		leap	pale	rap
ale	epi		Lear	paler	rape
alp	era		lepra	pali	re
ape	eria		lerp	par	Rea
April			liar	pare	real
ar			lie	parel	reap
are			lier	pea	rei
Ariel			lip	peal	rep
aril			lira	pear	rial
				pearl	rip
				per	ripe
				peril	rile
				pi	
				pia	
				pie	
				pier	
				pile	
				plea	
				plier	

Results.—(1) *Norms* are given in Tables 130 and 131 for the two forms of the test separately and so distributed as to show the average performance for each sex at each age. These norms have been compiled by combining the data obtained by Anderson and by Pyle. Percentile distributions for the scores of the two tests added together are shown in Figs. 77 and 78. These curves have been derived from Anderson's data by subjecting the raw data to the process of numerical smoothing and by further smoothing the curves in the process of drafting them.[1]

TABLE 130

Averages by Age and Sex, aeobmt-Test (After Anderson and Pyle)

SEX	AGE	8	9	10	11	12	13	14	15	16	17	18	ADULTS
Boys	Cases	38	86	103	112	124	134	109	85	59	49	37	70
	Aver.	5.9	9.0	8.7	10.5	11.4	11.5	13.2	15.3	19.3	17.0	16.9	18.4
Girls	Cases	48	101	128	115	141	99	118	96	94	70	53	93
	Aver.	7.5	8.0	10.8	11.8	13.2	14.9	15.0	15.8	15.8	15.9	18.4	20.9

[1]See Ch. 3, p. 31, for these methods.

On account of the relatively small number of cases available for each curve, its topography must be regarded as somewhat provisional, though the error is presumably within one or two points. It will be understood that the minimal and maximal scores do not afford permanent standards of comparison.

TABLE 131

Averages by Age and Sex, acirlp-Test (After Anderson and Pyle)

SEX	AGE	8	9	10	11	12	13	14	15	16	17	18	ADULTS
Boys	Cases	39	88	102	112	130	144	111	87	63	52	39	30
	Aver.	5.5	7.3	8.3	10.6	11.5	12.6	13.9	16.2	17.0	19.3	16.4	21.8
Girls	Cases	41	97	124	114	138	94	121	98	94	71	54	45
	Aver.	6.5	7.7	10.2	11.5	13.3	14.7	16.2	17.4	17.7	18.0	19.3	21.4

(2) *Dependence on age.* Mrs. Squire's conclusion that the correlation "between efficiency and maturity is not so complete as in many of the other tests" would seem to be based upon the examination of too few cases (10 of each age), for the averages of Tables 130 and 131, with the single exception of that for the 9-year-old boys in the *acobmt*-test, show a progressive rise with age from 8 to 17 years, while adults are uniformly superior to the boys and girls of 17. The same thing is brought out in the combined results set forth in the curves: that for 9-year boys lies for the most part above that for 10-year boys. There is also a drop in the upper percentiles of 15-year-old girls that is difficult of explanation unless there has been some accidental selection of poor *S*'s in this group.

(3) *Dependence on sex.* Comparison of the two tables and of the two charts makes it easily evident that girls, at least up to the age of puberty, are consistently superior to boys and by an amount approximately equal to one year's advance. It follows that this sex difference must always be had in mind in the use of the norms and distributions for this test.

(4) *Dependence on race.* Pyle's comparative study of whites and negroes (3a) shows clearly the inferiority of the latter in this test. In terms of general averages for all ages (in which there is, unfortunately, a certain element of unreliability owing

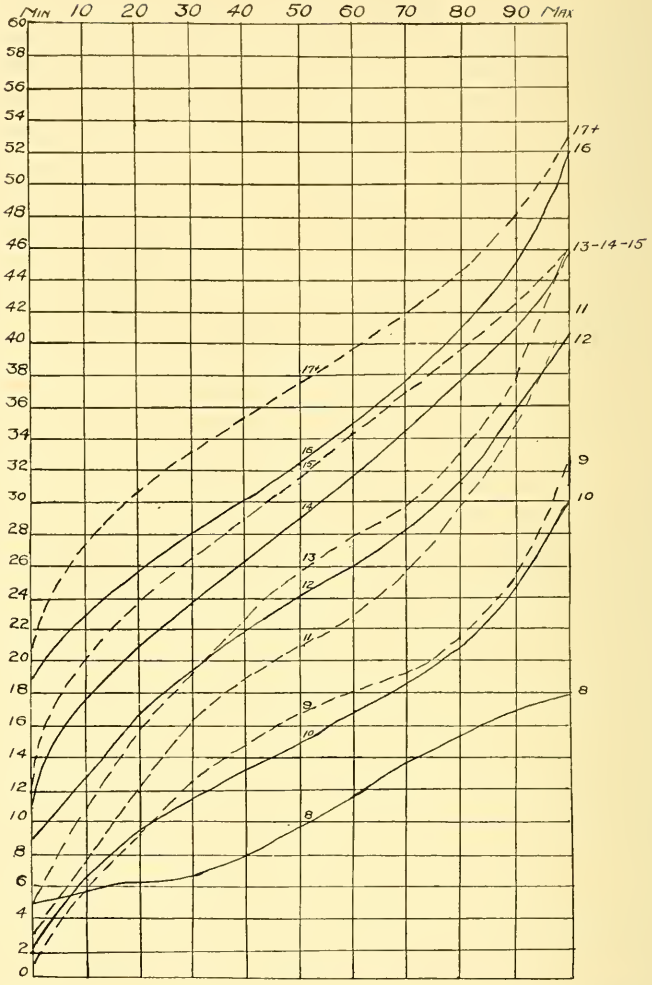

FIG. 77. PERCENTILES OF WORD BUILDING FOR BOYS. (Anderson)

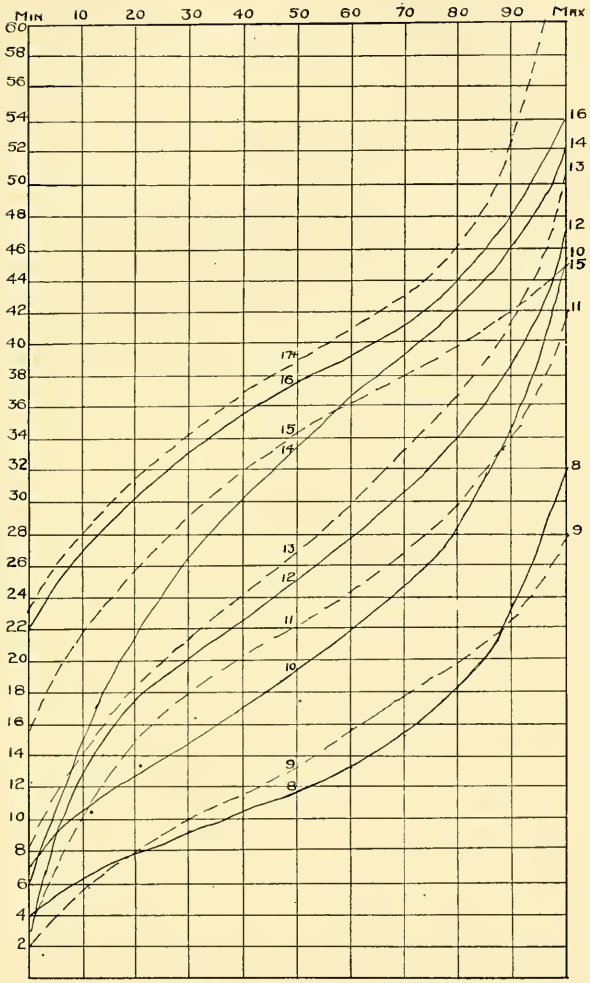

FIG. 78. PERCENTILES OF WORD BUILDING FOR GIRLS. (Anderson)

to the unequal numbers of the two races tested at different ages), the male whites scored 10.8 words in the *acobmt*-test and 11.3 words in the *acirlp*-test; the male negroes scored 5.2 and 6.0 words for these tests, respectively. Similarly, the female whites scored 12 and 13 words against 5.9 and 5.1 for the female negroes in these same two tests. These differences are, of course, well outside of their probable error.

(5) *Individual differences* are decidedly large in this test. Inspection of the charts, for example, will show that some *S*'s at 9 are superior to some *S*'s at 17 and over. Similarly, in the author's first trials of the *acobmt*-test, 10 of his 36 grammar-school *S*'s scored 15 words or over, while 13 of his college students scored fewer than 15 words. This wide-range distribution of the scores is an obvious point in favor of the use of the method in diagnosis of individual status.

(6) *Frequency of different words.* By examining the papers in detail, and tabulating the total number of words formed and the number of times each of these words is given, one may discern something of the principles which govern the operation of the test. The following are the data secured by the author:

TEST NO. 1. 58 COLLEGE STUDENTS. (43 DIFFERENT WORDS.)

Over 50 times—bat, mat, bet.
40-49 times—eat, met, Tom, at, boat.
30-39 times—meat, to, tea, beat, team, tab, ate, am, moat, mob, me, beam, toe.
20-29 times—tame, oat, be, mate.
10-19 times—boa, mote, bate, abet, tomb, tome.
5- 9 times—Mab, Abe, Mae, ma, atom, a.
1- 4 times—bot, mot, o, Moab, beta, bema.

TEST NO. 1. 50 GRAMMAR-GRADE BOYS. (38 DIFFERENT WORDS.)

Over 40 times—mat, bat.
30-39 times—bet, at, met.
20-29 times—to, eat, Tom, beat, tea, meat, be, am, boat.
10-19 times—toe, mob, beam, me, ate, team, tab, boa, oat.

5- 9 times—ma, bate, a, moat, mot, tame, mate, bot.
1- 4 times—tam, tomb, Abe, mote, Moab, Mae, o.
Not given—those not given by college students, plus abet, atom, bema, beta, Mab, tome.

TEST NO. 2. 69 COLLEGE STUDENTS. (59 DIFFERENT WORDS.)

Over 60 times—lip, lap.
50-59 times—rip, rap, pear, ear, real, pie, leap, rail, pale, reap.
40-49 times—pail, pile, ale, pair, are, ape, lie, pea, peal.
30-39 times—pare, earl, pearl, air, par, lair, ripe, liar.
20-29 times—ail, Lear, rape, ire, pal.
10-19 times—lea, pa, rile, pire, era, pier.
5- 9 times—per, a, alp, Eli, plea.
1- 4 times—I, paler, peril, lira, rep, rale, ra, April, Ira, la, pi, Rea, Rae.

Inspection of these lists shows (*a*) that three-letter words are in every instance those most frequently formed, (*b*) that two-letter words and the one-letter words, which one might expect to be most frequent since most simple, stand relatively low, *e. g., ma, be, am, pa, me, a, o, I,*[2] (*c*) that grammar-school boys give all the words given by college students save a few rather unusual terms such as *atom* and *tome,* (*d*) that usage and ordinary speaking vocabulary condition the formation of words, in as much as the most ordinary words have the greatest frequency, *e. g., bat, mat, bet, cat, lip, lap,* whereas words that are less frequently used in every-day speech, although their meaning is doubtless perfectly well known, do not suggest themselves so readily under the conditions of the test, *e. g., tomb, tome, era, plea, paler,* (*e*) that the words not given by any one are, with one or two exceptions, *e. g., plier,* words of extremely rare usage or unusual form, alternative spellings, etc.

(7) *Reliability.* Wyatt found a coefficient of reliability of .88 between the results for two different arrangements of the

[2]It appeared, upon inquiry, that some of the college students had omitted words like *pa, ma, a, o,* and *I* on the ground that they were 'not real words,' or 'didn't count,' but, oftener, they seem to have been passed over because the attention was concentrated upon the making of *combinations.*

letters *acobmt*. A sample group of 46 cases from Anderson's data gave a reliability of .74, P.E. .07, when the *acobmt*-test and the *cairlp*-test were compared. The word-building test is seen, therefore, to possess a good degree of reliability.

(8) *Correlations.* Heymans and Brugmans found positive correlations of from .12 to .76 between word-building and five other tests of imagination (puzzle picture, .35, solving riddles .24, arrangement of syllables .76, jig-saw puzzles .12, Binet's paper-cutting test .47). Wyatt's corrected correlations with word-building gave with analogies .93, with the completion test .97, with the part-wholes test .99. The raw correlations for his two groups of *S*'s were for analogies .54 and .65, for the completion test .36 and .70, and for the part-whole test .36 and .77. He also found raw correlations of .39 and .52 for a test of sentence-construction, a correlation of .47 with interpretation of fables, but no correlation with the letter-square test.

The author found no correlations between word-building and class standing in the case of grammar-school pupils and the insignificant correlation of .13, P.E., .08, in the case of 58 college students. Terman, however, found his stupid boys generally inferior to his bright boys.

(9) *Conditioning factors.* Age, sex and general intelligence are not the only factors that affect the outcome of this test. Thus Terman remarks: "Much depends, of course, upon the vocabulary at command, and this in turn depends largely upon home training and amount of habitual reading as well as upon native retentiveness. A second factor is ability to spell, and habits of word analysis generally. Very important, also, is the use of a rational plan; some skipped about and made combinations at random, while others took the letters one by one and joined them in as many different ways as possible with the others. Lastly, the rate of shifting of attention, and the degree of mental inertia as opposed to spontaneity, also contribute to the result" (5, p. 342).

REFERENCES

(1) E. J. Anderson, Standardization of some mental tests. (Study from the Cornell University Educational Laboratory, as yet unpublished.)

(2) G. Heymans und H. Brugmans, Intelligenzprüfungen mit Studierenden. *ZAngPs*, 7: 1913, 317-331.
(3) W. H. Pyle, The examination of school children. New York, 1913. Pp. 70. Especially 22-24.
(3a) W. H. Pyle, The mind of the negro child. *School and Society*, 1: 1915, 357-360.
(4) Carrie R. Squire, Graded mental tests. *JEdPs*, 3: 1912, 363-380, 430-443, 493-506, especially 441-442.
(5) L. M. Terman, Genius and stupidity: a study of some of the intellectual processes of seven 'bright' and seven 'stupid' boys. *PdSe*, 13: 1906, 307-373.
(6) G. M. Whipple, Vocabulary and word-building tests. *PsR*, 15: 1908, 94-105.
(7) S. Wyatt, The quantitative investigation of higher mental processes. *BrJPs*, 6: 1913, 109-133.

TEST 48

Ebbinghaus' completion method.—In July, 1905, the school authorities of Breslau requested certain persons, among them Professor H. Ebbinghaus, to undertake a scientific investigation of the fatigue-effects of the continuous five-hour session then in vogue in that city. In the course of this investigation Ebbinghaus devised and applied, in conjunction with other tests, what he termed the '*Combinationsmethode*' (since referred to by Elsenhans as the 'completion-method' and by others as the mutilated text or missing-words test).[1]

The author of the method says in substance: Mental ability demands not merely retentive capacity, readiness of recall, or facile association of specific past experiences; it demands all this and something more, something more complex and, as it were, creative; namely, the ability to combine, into a coherent and significant whole, mutually independent and even seemingly contradictory impressions. In short, intelligence is essentially a combinative activity. To measure intelligence, there-

[1]Meyer has pointed out the inaccuracy of the translation "combination-method," which has been current for some time. The German *Combinationsgabe* is not a talent for combination, but an ability to "put two and two together," or, to use Meyer's explanation, "a talent for drawing conclusions from premises which do not very readily present themselves to a man's consciousness as items of a unitary logical thought, but which, *as soon as they are combined*, suggest the conclusion very forcibly." This is quite true, but the author can not see that Meyer has improved matters by advocating the translation "conjectural method." To conjecture is to surmise, to guess, to form a tentative opinion, inferentially. Technically, the activity in the Ebbinghaus test might be labelled 'redintegration,' but, as this term is somewhat clumsy, the designation 'completion method' seems entirely adequate.

fore, we must employ a test that demands ability to combine fragments or isolated sections into a meaningful whole. Such a test may be afforded by mutilated prose, *i. e.,* by eliding letters, syllables, words, or even phrases, from a prose passage and requiring the examinee to restore the passage, if not to its exact original form, at least to a satisfactory equivalent of it.

On account of the enthusiastic statements of Ebbinghaus, who characterizes this method as "a real test of intelligence," and as "a simple, easily applied device for testing those intellectual activities that are fundamentally important and significant both in the school and in life," the test has assumed some prominence.

The classification of this method in a system of tests is not always easy, for the simple reason that what mental processes it demands depends almost entirely upon the number and kind of elisions that are made in the text. To take extreme cases, if the elisions are numerous and sweeping, it may become really a linguistic puzzle of a very difficult variety, and it then belongs rather in the group of tests of active or creative imagination of the literary type; if, on the other hand, the elisions are but few and simple, it may degenerate into a simple test of controlled association of any desired degree of ease. Again, if the original text be first read to the examinee, as some, *e. g.,* Elsenhans, suggest, the test becomes in the main a test of associative recall, *i. e.,* a form of memory test.

Since the elision of a single letter may, in some circumstances, very considerably increase the difficulty of the test, it follows that, without extensive preliminary trials, it is well-nigh impossible to prepare a series of texts of equivalent difficulty, or to insure that the several sections within a given text present equivalent difficulty.

That these difficulties in the preparation of the text are real and serious is attested by the unanimity with which they are expressed by all investigators. They have led some experimenters to question whether the method did, after all, get at the mental activity it was designed to call forth, but the trend of opinion has been on the whole distinctly and even enthusiastically in favor of the test.

The following is a sample section of text as used by Ebbinghaus and other German investigators: the dotted lines indicate the position and approximate length of the omissions.

Belagerung Kolbergs. 1807.

"Da der Feind fortf.....an....neuen Schanze am Sandwege.... angestr.....Eifer zu.....so hatte unser neuer Kommandant gleich.....ersten Nacht.....Hierseins einen Aus.....dieselbe angeordnet," etc.

Terman elided, in the main, whole words, instead of syllables, on the ground that the word is a more natural unit of language than the syllable, and that ability to supply missing syllables will, in the case of school children, depend largely on the extent to which word-analysis has been taught in the schools: this varies in different school systems and even in different classes of the same system.

MATERIALS.—Stop-watch, or for group work, the special seconds clock is recommended. Four printed texts. [If all four texts are to be used for the test, *E* should prepare a short sample piece of mutilated text, say three or four lines, which may be typewritten, or placed on the blackboard for group work, and used for demonstration and preliminary trial. If one of the texts is not used, this may serve the purpose.]

Text No. 1, prepared by the author, has been used by him in tests upon college students and by Mrs. Squire in tests upon school children. It contains 100 elisions, including some in which, in accordance with Ebbinghaus' plan, portions of words as well as entire words are elided.

Text No. 2, taken from Terman (18), has been used by Wyatt in tests upon English school children; it contains 93 elisions.

Text No. 3 is designed especially for use after a preliminary reading of the entire completed form. It is taken from Terman and contains 100 elisions.

Text No. 4, taken from Terman and Childs (19), and not here reproduced, is substantially the same passage as No. 2, but the elisions are made upon a new plan, such that there are four sections representing four different degrees of elision. In the first section 33, in the second 45, in the third 54 and in the fourth 66 per cent. of the original material is elided. This text demands a special system of scoring. In use in the author's laboratory it has been found that the second and third blanks are peculiarly unfortunate: they are too difficult and tend to

produce discouragement and confusion at the outset. Many adults also find the fourth section easier than the third, despite the greater amount of elision in it. However, the text has been retained as presented by its authors on account of the norms published for it by them.

The reader who desires to try yet other texts will find eight of them in Simpson (16, pp. 119-121).

In the reproductions here given, italics indicate the elisions in the printed text.

Text No. 1.

Where the Dandelions Went.

When Willy *was* two *years* old, he *lived in a* red farm-*house with a* yard *in* front *of it*. The dan*delions* were *very thick* there; so that *the yard* look*ed* yellow instead of *green*.

One bright *day* Willy's m*amma* put *on his* straw *hat and* sent him *out into the* yard to *play*. She knew *the yard* had *a* high *fence;* and *he* could not *open the* gate; so he *was safe*. *When* it *was* time *for* him *to have a* nap and *she* went *to call* him, she noticed that *a great many* of the dan*delions were gone*. She won*dered* where *they were;* but, as *Willy could* not talk much, *she* did not *ask* him *about* them.

A short *time* after, while *he* was *asleep in* his crib, *his* mamma went *out to draw some* wa*ter*. When *the* buck*et* came *up* full *of water*, *the* top was all *yellow* with *dandelions*. Look*ing* down into *the well*, *she* could *see* no *water* at all, on*ly dandelions*.

It was no wonder, then, where *the* blos*soms* had *gone*. Willy *had* been *very* busy, try*ing to* fill *up the well*.

Text No. 2.

The Strength of the Eagle.

One *day the* eagle *went* with the *other* birds *to* see *which* could *fly the* highest. *They* agreed *that* he who *could* fly *the highest* should *be* called *the* strongest *bird*. All started *at the* same *time* and *flew* away among *the* clouds. One by one they *grew* weary *and returned*, but *the* eagle *flew* upward and *upward* until *he* was *a* mere speck *in the* heav*ens*. When he *came* back, *the* others were *waiting* for him; and *when he* touched *the ground* a linnet *flew* off *his* back where *he had been* hidden and *said* that *he* himself *was the* strongest *bird*. "*I am* stronger *than the eagle*," said the *linnet*, "for not *only* did I *fly* as high, but *when* he began *his* downward *flight*, I *left my* hiding *place* and *flew* up *a* little *higher*." *At* this boastful *speech* the *others shook* their heads and *called a* council to *decide* the matter. After *a* long *debate they* decided *that* the *eagle was* the *strongest* bird, *for* not only *did* he *fly* so high, but *he carried* the *linnet* as well.

To *this* day *the* plumes *of the eagle* are emblems of *strength* and cour*age*.

Text No. 3.

Why the Mole is Blind.

An *Indian* once *chased a squirrel* into cloud*land*. Then *he set a* trap *for him*, laughing to *think* how *he* would *catch him*. The *squirrel* did

not come back, but *alas!* the *sun* on *his* daily *rounds fell* right *into the trap.*

When *the* bright *sunlight did not* come, the *Indian* began *to* be *uneasy,* and when *he* found *his trap* had *the sun* fast he *did not* know *what to* do.

He tried *to get* near enough *to loosen* the cords, but *the heat* from the *sun scorched* him and he gave *it up.*

Then *he coaxed* many *animals* to try *it,* but *they* all found *the sun* too *hot.* At *last* the *mole said:* "I will *dig* through *the ground* under the *trap* and so get at *the cords.*"

This *he did* and the *sun leaped* up *into the heavens.*

But it *went* so *quickly* that *the* poor *mole could* not *get away,* and the *heat* of *the sun put* out *his eyes.*

Since then the *moles* have had *to live in* dark *places,* and unless one *looks* very *closely* he *can not* find *their eyes.*

METHOD.—Provide *S* with a demonstration or practise text (either one of the three regular texts not to be used subsequently—except that Text 2 should not be used if 4 is to follow—or the special sample prepared by *E*). Explain the nature of the test, in accordance with the directions printed on the test-blanks. It is well, in addition, to suggest that, in case a certain elision offers special difficulty, it may be temporarily passed by, since the correct interpretation of the context further on will often give the necessary cue for the omitted elision.

When it is clear that *S* understands the conditions, proceed with the test proper. If but one trial is to be made, use Text 2 or 4 with a 10-min. limit. If more than one, follow with Text 1, using the work-limit method.

Record the time and make notes of the manner in which *S* undertakes the test. Does he read it all over first? Does he work systematically? Attentively? With confidence or hesitation? Does he grasp the general thread of the story?

VARIATIONS OF METHOD.—(1) To conduct the test with the memory feature, employ Text No. 3, which is specially devised for that purpose. After the preliminary trial, read the unmutilated text for No. 3, entire, to *S*. Then supply him with the No. 3 test-blank and proceed as before. The text may be read more than once, or any desired time-interval may be introduced between the reading and the execution of the completion. Other variations will suggest themselves, *e. g.,* auditory, visual, or auditory-visual reading, etc.

(2) To approximate the conditions observed by Brown, Wyatt and others, give *S* opportunity (say three to five minutes) to examine the text before filling it out. This variation of method obviously changes the character of the test considerably: it tends to greater uniformity in the mental processes of the *S*'s, but it removes the differentiation which the standard method conserves in that some *S*'s are quick to see the necessity of looking over the text ahead of their work while others are not.

(3) The Ebbinghaus test lends itself rather well to group tests. With Texts 1, 2 and 3 the use of the author's seconds-clock[2] is recommended. If, however, the time-limit method is followed, the limit must be so chosen that the fastest *S* in the groups under comparison can but just complete the work. For adults, 7 min. may be employed for Texts 1 and 2, a shorter time for Text 3. Text 4, it should be noted, is devised to be scored by the time-limit method only. For it Terman and Childs specify 15 min. Unfortunately, this time is too long for some high-school students, as investigations in the author's laboratory have shown; in fact, even when the time is shortened to 10 min., a few pupils (about 2 per cent.) will finish before that time is up.

Treatment of Data.—Text 4, as already mentioned, is used with a constant time-limit (15 min., according to its originators, but preferably 10 min., according to the author's experience). It is scored by assigning for each correctly filled blank 6 units in Section I, 8 units in Section II, 10 units in Section III and 13 units in Section IV. The total score is then divided by 10 and amounts to 100 (exactly, 100.2). One-half the above credits are given if the inserted words "make a well-connected story," but are "related in only a moderate degree to the thought that should have been given." No credit is given for inserted words that make no sense in their setting nor for words that make a continued story which is "purely literary invention, having no connection with the thought given by the printed words." Thought is "considered rather than elegance in diction."[3]

[2]See Vol. I, p. 9.
[3]For samples, see Terman and Childs, pp. 201-202.

For Texts 1, 2 and 3 there are three possible methods of scoring. For adults, working individually with these texts, the quality of the work is usually so good that speed alone may be used as an index of efficiency.

Secondly, speed may be neglected and attention given only to quality (in which case the instructions should be modified to indicate that *S* may 'take his time'). An example of this method may be seen in the work of Burt, who graded the worth of each inserted word on a system of 6 points, 0 to 5.

Thirdly, speed may be combined with quality, and in one of two ways. The quality may be determined and related to speed by means of formulas like those developed for the cancellation test (No. 26). Or, again, the time-limit method may be followed and the work scored by the plan proposed by Ebbinghaus himself and used by Krueger and Spearman, Brown, Wyatt and others. Here quantity and quality of work are computed as follows: (1) Give a credit of 1.0 for each elision filled in in any manner. (2) Give a debit of 0.5 for each elision unfilled in any manner. (3) Give a debit of 1.0 for each elision filled in such a manner as not to make sense, or for each word introduced in excess of the number called for by the lines that indicate elisions (or, if desired, also for each word that is quite obviously too short or to long for the space assigned for completion, even though the passage 'makes sense').[4] For quantity of work done, add (2) and (3) and subtract the sum from (1). For quality of work done, compute the relation in per cent. of the same sum to (1).

TYPICAL RESULTS.—The following is a sample of the work of a boy, 11 years old, one of Terman's "bright" group, who 'completed' Text 2, with the exception of three elisions, in 26 minutes. He was quick, steady, and looked ahead.

"One day an eagle went with the other birds to see who could fly the highest . . . (Next three sentences correct) . . . When he came back the others were waiting for him; and when he touched the ground a linnet flew off his back where the thief had hidden and said that he himself was the strongest bird. "I am stronger than you are," said the linnet, "for not alone did I fly as high, but as he began flying downward, then I left my hiding place and flew up a little higher," etc.

[4] Cohn and Dieffenbacher penalized only 0.5 for errors in the length of the inserted word.

The following is a sample of the work of a boy of the same age, one of Terman's "stupid" group, who worked for 25 minutes at the same text. Save in one or two easy sections, his 'completions' make no sense at all. He worked by phrases only.

"One with the eagle and with the small birds and see who could fly the highest, and agreed and he who will fly the highest should be called the strongest they All started in the same place and whent away among the clouds. . . . After a while he decided that the king of the little bird and not only and not he was so high, but he did the thing as well," etc.

Results.—(1) *Norms and dependence on age.* The results gained by Terman and Childs with Text 4 are shown in Table 132, in which the last column is presented as a basis for the use of the test diagnostically; *i. e.*, the score reached by 66 per cent. of the children of a given age is taken as the limiting standard of efficiency for that age. Tables 133 and 134 show the results

TABLE 132

Completion Text No. 4. 15-Minute Limit (Terman and Childs)

AGE	CASES	MEDIAN	P. E.	REACHED BY 66%
9	32	18.4	8.5	14.9
10	39	29.2	11.8	20.4
11	52	32.2	11.1	25.2
12	56	34.2	11.7	25.6
13	57	45.9	15.2	36.6
14	33	48.5	9.4	42.8

for boys and for girls, respectively, obtained by Mr. Fraser, of the Cornell Laboratory, with Text 4, but with a 10-minute time-limit. In these two tables the number of cases for every group is given in parenthesis just above the group average, and the groups are sorted to differentiate age and school grade as well as sex.

It will be seen that at 13 and 14 (when the sexes are combined), the Ithaca children have the better scores despite the fact that they worked but 10 min. This difference is partly due to the inclusion of high-school pupils in these ages, whereas the table of Terman and Childs is limited to children from the 4th to the 8th grades at Palo Alto, California.

Cohn and Dieffenbacher's tests show progress with age, even up to the 20th year.

(2) *Dependence on school training.* Wiersma called attention to the fact that the relation between performance in this test and age is far less evident than that between performance and scholastic status. This fact is illustrated clearly in the tables prepared by Fraser: taking, for instance, the results for girls, averages by ages, 13-19, run 70.9, 63.2, 57.3, 61.8, 65.8, 63.2, 70.4, and show no definite correspondence, whereas the averages by school grades, 8th to fourth year in the high school, run 38.9, 56.5, 67.7, 64.5, 80.7. Since there exists a distinct positive correlation between standing in this test and general intelligence, and since the higher the grade of pupils of a given age, the more intelligent, on the whole, they must be, it follows

TABLE 133

Completion Text No. 4. 10-Minute Limit. Results for Boys (Fraser)

AGE	GRADES					
	8	H. S. I	H. S. II	H. S. III	H. S. IV	All
13	(7) 46.3	(6) 54.1	(2) 32.6			(15) 46.9
14	(9) 36.5	(18) 54.1	(9) 46.3	(1) 54.6		(37) 48.2
15	(7) 39.5	(22) 48.5	(7) 61.9	(3) 44.1	(1) 92.6	(40) 50.0
16	(2) 42.1	(5) 38.7	(7) 55.9	(11) 73.3	(3) 73.9	(28) 60.6
17	(1) 34.0	(3) 42.5	(11) 61.5	(13) 58.0	(8) 62.5	(36) 58.1
18		(1) 33.2	(4) 51.1	(8) 43.7	(7) 56.5	(20) 49.1
19			(1) 32.0		(8) 61.9	(9) 58.6
All	(26) 40.3	(55) 49.4	(41) 54.1	(36) 58.2	(27) 63.2	(185) 52.9

TABLE 134

Completion Test No. 4. 10-Minute Limit. Results for Girls (Fraser)

AGE	GRADES					
	8	H. S. I	H. S. II	H. S. III	H. S. IV	All
13	(4) 66.3	(4) 75.3				(8) 70.9
14	(5) 37.3	(26) 62.4	(7) 83.2	(1) 71.0		(39) 63.2
15	(8) 46.9	(28) 50.4	(11) 74.1	(6) 72.7		(53) 57.3
16	(2) 34.4	(6) 45.0	(19) 66.9	(13) 61.8	(2) 89.9	(42) 61.8
17		(1) 49.0	(10) 57.3	(17) 65.7	(5) 86.6	(33) 65.8
18		(3) 54.3	(2) 30.0	(12) 63.9	(8) 75.1	(25) 63.2
19		(1) 79.5	(1) 82.0	(2) 46.4	(2) 84.2	(6) 70.4
All	(19) 38.9	(69) 56.5	(50) 67.7	(51) 64.5	(17) 80.7	(206) 61.1

that some part of this correspondence between school status and the results is due to the correlation with intelligence. The effect of the school training itself undoubtedly contributes another portion of the correspondence—just how much cannot be said.

(3) *Dependence on sex.* Although Wiersma could not make out sex differences with certainty, the subsequent work of Burt and of Fraser leaves little doubt that girls are superior to boys in this, as in most tests with verbal material. Burt found girls distinctly better than boys in his text *The Two Matches* (means 84 and 70, respectively) and slightly better in another test of a more argumentative character (53.2 to 50.4). Inspection of Fraser's tables will show that, with the exception of the 8th-year averages, the girls excel in every group,

whether comparison is made by ages or by school grades. The superiority of boys found by Burt and Moore in one test is attributed by them to the nature of the text, which was such as to appeal more strongly to the interests and knowledge of the boys. It is difficult to reconcile, however, the results reported by Cohn and Dieffenbacher, who found girls inferior to boys at all grades and regardless of equivalence in ages. These sex-differences were greater in the upper than in the lower classes and sufficient to bring the better girls on a level with the poorer boys of their age and grade. It is possible that this striking opposition to the results found by others may be due to some differences in the organization of the schools at Freiburg.

It is unfortunate that Terman and Childs have made no distinction between the sexes in reporting their averages and establishing their age standards.

(4) *Individual differences.* Distribution of the data obtained from the completion method reveals large individual differences. This is demonstrated by the large size of the P.E. in the data of Terman and Childs and even more definitely by the percentile curves of distribution for each sex prepared from Fraser's combined results for pupils from 14 to 17 years of age (Fig. 79).

(5) *Practise,* according to Wiersma, may effect an improvement in efficiency in the completion test that may be easily discerned after the lapse of 10 days, and even after an interval of 6 weeks. As a consequence, it is evident that, in making use of this test for comparative work at different periods, steps must be taken to eliminate or compute the practise-effect.

(6) *Dependence on fatigue.* In Ebbinghaus' Breslau investigation no fatigue effects could be made out as the result of the five-hour session in the tests with the upper classes, or at least, if fatigue were present, it was masked by practise. In the lower classes (10-12 years) there appeared to be a decrease in the quantity and quality of work toward the end of the session.

These conclusions have been criticized by several experimenters. Binet and Henri contend that the several texts were of too unequal difficulty and that the method of scoring was arbitrary and crude. Lobsien has also criticized the general

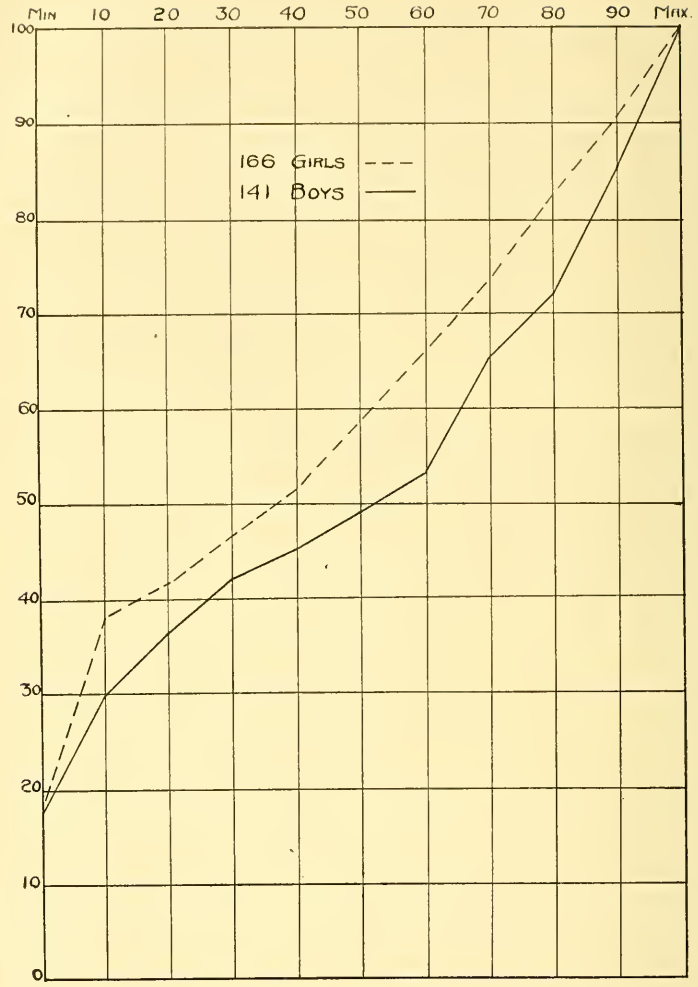

FIG. 79. PERCENTILES FOR COMPLETION TEST NO. 4. AGES 14 TO 17 COMBINED (Fraser).

plan of administration of the tests in the Breslau investigation. It is admitted that the material used was too easy for the upper classes, and that this circumstance tended to obscure the influence of fatigue and other factors.

In the opinion of Kraepelin, the Ebbinghaus test is to be regarded more as a device for exploration than as a decisive and accurate device for measuring fatigue, for, in the first place, no systematic study has yet been made of the relations between mental fatigue and the complex activities concerned in this test, and secondly, the evaluation of the errors is so difficult and their scoring so arbitrary that the test is not well designed for single applications (*Stichprobe*) and statistical treatment.

Wiersma compared performance before and after a 10-days vacation, but he expresses his belief that the marked improvement cited above as exhibited by the pupils at the second test was largely due to practice. It is unfortunate that proper measures have been not taken to eliminate the practice error in these, and in other applications of the completion test.

(7) *Dependence on intelligence.* It has already been intimated that the relation demonstrated between performance in the completion test and scholastic standing is partially conditioned by a direct correlation with intelligence. Such a relationship was found by Ebbinghaus, most clearly in the lower and progressively less clearly in the higher grades, by sectioning the Breslau pupils into three groups—best, average, poorest—on the basis of their class standing: these three sections scored 56, 48 and 43, respectively, in quantity of work, and 17.3, 20.8 and 26.3, respectively in quality of work (percentage of errors). Similar results were reached by Cohn and Dieffenbacher. Wiersma found a positive correlation between capacity in this test and native ability (*Begabung*), both in tests at a teachers' seminary (ages 14.5 to 19.5) and at a continuation school (ages 12 to 15).

Since then, statistical treatment by the more accurate correlation methods has continued to reveal positive correlations of good magnitude between the completion test and intelligence. Brown, for example, found a correlation of .43 with one group (66 boys, aged 11-12 years) and of .69 with another group (39 girls, aged 11-12 years), and he declares that the Ebbinghaus test "is a good measure of intellectual ability. It correlates with 'general intelligence' almost as closely as 'scholastic in-

telligence' (school marks) does." Burt reports correlations of .48 and .53, Wyatt of .85, P.E. .04, with one group and of .61, P.E. .07, with another group. Simpson, who compared the capacities of two contrasting groups of adults in numerous tests, found that the Ebbinghaus method almost completely separated his two groups, and that there was a correlation of .89 between results with his 'good' group and the estimated intelligence of the members of that group alone.

Wyatt applied the analogies and the completion tests to seven children who were candidates for admission to the Fielden School, Manchester, England, and was able not only to advise which of the candidates should be admitted and which not, but also to predict successfully the approximate position that these pupils would take in their class at the end of the term.

Ebbinghaus believed that the correlation between the completion test and school ability might become obscured (1) because the test puts a premium upon speed of work, whereas the school grade is based on work that permits of a slower pace; (2) because, in some part, standing in the test might depend upon purely formal linguistic skill or verbal dexterity—a form of ability which he thought had but a limited scope in school work; and (3) because the text selected for the test might be too easy.

In the author's opinion, these reservations are scarcely in order, in so far as Ebbinghaus implies that school grades are inferior to his test as a measure of intellectual ability and asserts that linguistic readiness plays no part in the determination of school grades.

The author is inclined, rather, to agree with Terman when he says: "My experience with the test causes me to regard it favorably; but, like all others, if taken alone, it can give only a partial account of the subjects' ability. It certainly does indicate something as to the general command of language. I am inclined to think that somewhat mechanical activities like memory and association, as distinguished from synthetic or combinative processes, play a relatively more important rôle in this test than Ebbinghaus assigns to them. Indeed, verbal memory, in the broad sense, would seem to be the chief factor in success." Incidentally, ability to spell, degree of familiarity with the type of literature from which the selection is taken, and the way in which *S* happens to go about the test may all affect his rank. Indeed, it is possible that a very original *S*, one with a spark of literary invention, might fare relatively poorly.

Terman and Childs say: "We believe that it [the completion test] brings to light fundamental differences in the thought processes."

(8) *Delinquents.* Text 4 has been used by the author, together with numerous other tests, in examinations of the mental status of certain selected 'citizens' of the George Junior Republic, carried on with the assistance of Mr. Fraser, at

Freeville, N. Y. While our data (Table 135) are too few to generalize from, they reveal, as far as they go, distinct reductions from the normal performances for *S*'s of the ages concerned, and these reductions, it is of interest to note, bring the averages in fair agreement with the standards corresponding to the *mental* ages at which these *S*'s had been rated by the other tests. These results, then, tend to confirm the conviction that the completion method is of considerable value in diagnosis of mental status.

TABLE 135

Completion Text 4. 15-Minute Limit. Results from 'Citizens' of the George Junior Republic (Whipple and Fraser)

SEX	CASES	CHRON. AGE	MENTAL AGE	AVERAGE	M. V.
Boys	13	14–19	10.0–12.5	27.6	13.6
Girls	5	15–18	9.6–11.2	35.7	15.4

(9) *Reliability.* Coefficients of reliability computed by different investigators for varying forms of this test have as a rule been quite high; thus, Brown finds the coefficient mostly over .70, Simpson .92 to .96, Burt .68, Burt and Moore, .58, and Wyatt .89.

(10) *Correlations with other tests.* Heymans and Brugmans report the following correlations with the Ebbinghaus test: discrimination of abstract terms .54, memory for ideas .56, problem solving .56, detection of grammatical errors .72. Wyatt found correlations as follows: with analogies .85, with word-building .70, with part-wholes .75, with interpretation of fables .69, with *anos*-test. 43, with memory for nonsense syllables .61, with dissected pictures .41, with letter-squares zero. For numerous correlations discovered by Brown with six different groups of *S*'s, consult the original texts (2b, pp. 122-123, or 2a, p. 316, or the same material may be gathered from Simpson, pp. 107-8). Simpson found correlations of .85 with hard opposites, .72 with easy opposites, .82 with memory for words, .71 with memory for ideas, .65 with adding, and .54 with the

g-test. Corrected correlations reported by Krueger and Spearman are as follows: completion test and pitch discrimination 0.81, completion test and adding 0.93, completion test and the hypothetical 'central-factor' 0.97. The completion test was not found to correlate with a test of memory span (*Auswendiglernen*). The extremely high correlation with the 'central-factor' is of special interest, since, if the argument be admitted, it demonstrates a very close dependence of performance in this test upon a certain hypothetical psychophysical capacity, presumably akin to plasticity of the central nervous system, which, in the opinion of these authors, is, for each individual, a fundamental conditioning factor in the performance of various forms of mental activity.

Notes.—At the risk of repetition, it may be pointed out again that the outcome of the completion test hinges largely upon the degree of difficulty of the text employed: too difficult or too easy texts are alike undesirable, for the former convert the test into a blind puzzle, while the latter fail to bring out characteristic individual differences of ability.

To use the test on an extensive scale, therefore, we need to have at hand a number of texts that have been standardized by comprehensive trials with groups of *S*'s of both sexes, various ages, and various degrees of capacity and training. In other words, we need a series of norms of performance, or 'coefficients of difficulty,' as it were, for an adequate number of prescribed texts. Tables 133 and 134 represent a contribution in this direction from the Cornell Laboratory. Any investigator can improve them by adding to them his own data.

The difficulty of making comparisons between the results of different texts applied at different times may be further reduced by always permitting each *S* to finish each text, and by distributing the texts to be compared in such a manner as to eliminate by subsequent computation whatever error arises from this difference of material.

The Lipmann-Wertheimer modification of the completion method is essentially as follows: a test-story is read to *S* to supply him with certain information which he is supposed thereafter to conceal. He is subsequently given for completion

a mutilated text, the elisions of which are so arranged as to trap him into introducing facts from the test-story which he is trying to conceal.

In an endeavor to retain the essential psychological features of the Ebbinghaus test and at the same time avoid the disturbance due to dependence upon linguistic aptitude, Healy has devised a *pictorial completion test.* This is carried on with the aid of a brightly colored picture, 10×14 inches, which represents an outdoor scene with ten discrete, simple activities going on. Ten one-inch squares are so cut out from this picture as to remove 10 different objects, each of which is essential to complete one of the activities. *S* is given the incomplete picture, the 10 cut-out portions, and 35 other one-inch squares of which 30 bear other objects while 5 are blank. His task is to insert the 10 squares that he judges essential to complete the picture. Data thus far published indicate that the test is difficult below the age of 9, that performance may be as good at 10 as at 13, and that some adults make poorer scores than children (due to their more critical attitude toward the drawing). A time longer than 5 min. with more than one 'illogical' or more than two 'total' errors is suspicious of defective mental ability in *S*'s above the age of 10. This test would appear to possess many possibilities of development.

REFERENCES

(1) A. Binet et V. Henri, La fatigue intellectuelle. Paris, 1898, especially Ch. 7.

(2) W. Brown, (*a*) Some experimental results in the correlation of mental abilities. *BrJPs*, 3: 1910, 296-322. (*b*) Same material, with slight modifications, appears as Ch. 3 in The essentials of mental measurement. Cambridge, Eng., 1911.

(3) C. Burt, Experimental tests of higher mental processes. *JEPd*, 1: 1911, 93-112.

(4) C. Burt and R. C. Moore, The mental differences between the sexes. *JEPd*, 1: 1912, 273-284, 355-388.

(5) J. Cohn und J. Dieffenbacher, Untersuchungen über Geschlechts-Alters-und Begabungs-Unterscheide bei Schülern. *Beihefte zur ZAngPs*, 2: 1911. Pp. 214, especially 30-36.

(6) H. Ebbinghaus, Ueber eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern. *ZPs*, 13: 1897, 401-459.

(7) T. Elsenhans, Nachtrag zu Ebbinghaus' 'Kombinationsmethode.' *ZPs*, 13; April, 1897, 460-3.

(8) D. K. Fraser, Unpublished investigation in the Cornell Educational Laboratory on the standardization of mental tests for the period of adolescence.
(9) W. Healy, A pictorial completion test. *PsR*, 21: 1914, 189-203.
(10) G. Heymans und H. Brugmans, Intelligenzprüfungen bei Studierenden. *ZAngPs*, 7: 1913, 317-331.
(11) E. Kraepelin, Ueber Ermüdungsmessungen. *ArGsPs*, 1: 1903, 9-30, especially 17.
(12) F. Krueger und C. Spearman, Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten. *ZPs*, 44: 1907, 50-114.
(13) O. Lipmann und M. Wertheimer, Tatbestandsdiagnostische Kombinations-versuche. *ZAngPs*, 1: 1907, 119-128.
(14) M. Lobsien, Ueber die psychologisch-pädagogischen Methoden zur Erforschung der geistigen Ermüdung. *ZPdPs*, 2: 1900, 273-286, 352-367.
(15) M. Meyer, An English equivalent of "Combinationsmethode." *JPh*, 9: 1909, 688.
(16) B. R. Simpson, Correlations of mental abilities. *Columbia ConEd*, No. 53, New York, 1912. Pp. 122.
(17) Carrie R. Squire, Graded mental tests. *JEdPs*, 3: 1912, 363-380, 430-443, 493-506, especially 442-443.
(18) L. Terman, Genius and stupidity. *PdSc*, 13: 1906, 307-373, especially 342-7.
(19) L. M. Terman and H. G. Childs, A tentative revision and extension of the Binet-Simon measuring scale of intelligence. *JEdPs*, 3: 1912, 61-74, 133-143, 198-208, 277-289, especially 198-202.
(20) E. Wiersma, Die Ebbinghaus'sche Combinationsmethode. *ZPs*, 30: 1902, 196-222.
(21) S. Wyatt, The quantitative investigation of higher mental processes. *BrJPs*, 6: 1913, 109-133.

TEST 49

Interpretation of fables.—In 1903 Swift suggested that mental ability might be measured by determining the capacity to interpret the typical situation given in a typical fable. The three test-fables employed by Swift were later (1906) used by Terman in his comparative study of bright and stupid boys. Still later (1912) Terman and Childs published a set of eight fables selected by trial from a series of 20, with the idea that responses to this test would assist in mental diagnosis in conjunction with the Binet-Simon and other tests. These authors believe that this 'generalization test,' as they term it, "will prove a usable addition to the scale. It presents for interpretation situations which are closely paralleled in human social relations. It tests the power to unravel the motives underlying acts and attitudes, to look behind the deed for the idea that prompted it. It gives a clue to the status of social conscious-

ness. This, if correct, is tremendously important for the diagnosis of the upper range of mental defectiveness." . . . "It does not need to be unduly complicated by language difficulties, as is always the case to greater or less degree in tests of ability to interpret poetry." [1]

MATERIALS.—Printed sheets, containing the 8 fables selected by Terman and Childs.

METHOD.—Instruct *S* substantially as follows: "I am going to read to you some fables. A fable is a little story that is meant to teach a lesson (convey a moral). After I have read each fable, I want you to tell what is the lesson that you think it teaches (the moral that it is intended to convey)." Read each fable twice through before asking for its point. For ordinary testing use Fables I, III, VII and VIII. These four have been specially selected by Terman and Childs from the eight supplied, as proving in actual test to answer best the expected requirements of due progress in scores with advance in age.[2]

It is best to conduct the test individually and to record verbatim the replies given orally by *S;* it is possible then to follow up *S*'s response by a few discrete questions if necessary to be positive of his interpretation. If group tests are made, supply each *S* with a blank sheet containing eight numbered spaces in which the replies may be written. The norms of Terman and Childs that follow are based upon such group tests with written replies for Fables I, III, VII and VIII.

VARIATIONS OF METHOD.—If time permits, supplementary or control tests may be made by the use of the four remaining fables.

TREATMENT OF DATA.—We quote as follows from Terman and Childs: "The difficulty of finding a method of scoring which does not give too large play to the personal equation is a serious criticism of the fables test. After experimenting with a number of methods the following system was adopted as the one best suited to bring out objective differences and to call atten-

[1] See, for example, the test used by Bonser, p. 8.
[2] See Terman and Childs, p. 138, for the criteria upon which these four fables were finally chosen.

tion to certain types of answers significant for clinical purposes:

"(*a*) A completely generalized and entirely relevant reply, 5 units.

"(*b*) A generalization, quite plausible, but slightly differing from the correct one, or else a correct statement mostly generalized but not perfectly free from the concrete, 4 units.

"(*c*) Correct idea stated in purely concrete terms, 3 units.

"(*d*) An irrelevant generalization, 2 units.

"(*e*) A reply in concrete terms with just a trace of relevancy, 1 unit.

"(*f*) No response, or an entirely irrelevant concrete statement, 0."

"Elegance, grammatical correctness, spelling, etc., should have no weight in the scoring. On the other hand, it is necessary to be discriminating as to essential thought in the response. The tendency of the inexperienced scorer is to give too much credit. "In practise there is a tendency to make sparing use of Scores 1 and 4, reserving 1 for a few replies that are not quite bad enough for 0, and 4 for a few replies, which, though pertinent and generalized, are not quite what is wanted.[3]

The following samples of scoring for replies to the four standard fables will serve as useful illustrations:

Fable I. The Maid and the Eggs.

Score 0. "She wanted to be dressed nice and be praised."
Score 2. "Not to carry things on the head." "Not to be selfish." "Not to boast."
Score 3. "If the maid had not planned so far ahead she would not have dropped her milk." "Don't make schemes for the future while you are carrying milk."
Score 5. "Don't count your chickens before they are hatched." "Not to build air-castles." "Don't plan too far ahead."

Fable III. Hercules and the Wagoner.

Score 0. "Hercules was not kind." "Hercules was selfish."
Score 2. "Teaches politeness." "Teaches not to be mean." "To do as you are told."
Score 3. "The lazy man should get out and try to push the wagon out himself." "When you get stuck in the mud, don't call for help, but try to get out yourself."
Score 5. "God helps them who help themselves." "Teaches us to help ourselves before we ask others to help us." "Don't depend upon others."

[3]For further discussion and samples of scoring, see Terman and Childs, pp. 135-139.

Fable VII. The Fox and the Crow.

Score 0. "The fox wanted the piece of meat." "The crow ought not to have tried to sing till she had swallowed it."
Score 2. "Not to be stingy." "Not to steal." "Think before you act."
Score 3. "The crow was flattered by this speech." "The crow was too proud of her voice." "If the crow had not been so flattered, she would not have lost her meat."
Score 5. "Do not let people flatter you." "Don't listen to praise."

Fable VIII. The Farmer and the Stork.

Score 0. "The farmer ought to have let the stork go." "The farmer was a bad-tempered man."
Score 2. "To be merciful." "Do not kill animals." "Don't blame the other fellow." "Never go into traps." "Not to tell lies." "Take what you get without squealing."
Score 3. "The stork should not be caught with bad people like cranes." "The stork was caught in bad company and had to be treated the same."
Score 5. "Keep out of bad company." "You are judged by the company you keep."

RESULTS.—(1) *Norms.* The results obtained by Fables I, III, VII and VIII applied to about 350 pupils from the 4th to the 8th grades, inclusive, in 14 rooms of the Palo Alto and Mayfield, Cal., schools provide representative data for children up to 13 years of age. Table 136 shows the percentage of children of a given age that secured each of the scores from 0 to 5 for each of the four fables. Experimenters can add their own results to this distribution and thus increase the reliability of the distributions. The data in Table 137 are secured by adding together the scores of the four fables (maximal score = 20) and multiplying by 5 to bring to a percentage basis. The last column of this table may be taken as a diagnostic basis, as it indicates the score that is reached by two-thirds of the pupils of a given age.

(2) *Dependence on age.* Terman believes that "what is tested by the interpretation of fables is, in part at least, that general change of mental horizon that comes with increased experience and dawning maturity." The data thus far available indicate a fairly steady increase of proficiency with age, save that performance at 12 does not differ greatly from that at 11 years. With these four fables at least there is evidently not much chance of obtaining distributions for ages less than 9; indeed, in the opinion of Terman and Childs it would be difficult to

TABLE 136

Distribution by Percentages in Interpretation of Fables (Terman and Childs)

FABLE	AGE	CASES	SCORES						TOTAL	PERCENTAGE OF INCORRECT GENERALIZATIONS
			0	1	2	3	4	5	4 + 5	
I.	9	41	14	31	24	4	4	19	23	51
	10	53	13	22	20	4	9	30	39	34
The Maid	11	61	8	6	31	10	16	28	44	41
and	12	80	7	10	22	5	16	39	55	28
the Eggs.	13	73	5	4	18	8	15	47	62	22
	14	43	5	9	20	5	16	38	54	27
III.	9	41	14	7	19	29	12	16	28	40
	10	53	9	9	13	19	17	32	49	25
Hercules	11	61	5	6	15	13	20	41	61	22
and the	12	80	3	12	12	11	21	39	60	17
Wagoner.	13	73	0	5	15	12	18	47	65	18
	14	43	5	0	9	2	36	45	81	10
VII.	9	41	19	27	34	14	2	2	4	90
	10	53	17	26	35	2	5	13	18	66
The Fox	11	61	8	15	44	8	8	16	24	65
and	12	80	11	17	32	6	15	17	32	50
The Crow.	13	73	11	16	27	3	8	32	40	40
	14	43	14	7	34	5	14	25	39	46
VIII.	9	41	36	19	31	5	2	5	7	81
	10	53	21	24	24	2	4	24	28	46
The Farmer	11	61	16	11	21	6	11	33	44	32
and	12	80	15	15	22	5	7	33	40	35
the Stork.	13	73	8	8	14	8	14	46	60	19
	14	43	7	5	18	5	11	52	63	22

The percentage of incorrect generalizations is obtained by dividing the number of Scores 2 by the combined number of Scores 2, 4 and 5 (see explanation of scoring given above).

select any fables that would be serviceable for younger children. On the other hand, the addition of a few more difficult fables "would make the test especially valuable at the upper end of the scale and help a great deal in the difficult task of extending the scale beyond 13 years."

(3) *Dependence on intelligence.* Swift reports that the table-test showed no superiority for the 'bright' children; that, on

TABLE 137

Norms for Interpretation of Four Fables (Terman and Childs)

AGE	CASES	MEDIAN	REACHED BY 66 %
9	41	35–40	25–30
10	53	45	35–40
11	61	50–55	45–50
12	80	55	45–50
13	73	70	55–60
14	43	70–75	55–60

the contrary, with the fable of *The Fishes and the Pike*, which, he says, requires "a distinctly intellectual process," the dull group excelled the bright group, while the answers of boys in an industrial (reform) school "distinctly outranked those from both of the other groups, but especially the ones from the 'bright' division, in the penetration and versatility that they showed."

Terman's tests showed, on the contrary, that his 'dull' group was distinctly inferior to his 'bright' group, for, as he says: "in the first place, they more frequently miss the point of the story altogether," and "in the second place, the dull boys are plainly deficient in degree of abstraction. Even when they give an approximately correct interpretation, they usually express it in the concrete terms of the given situation, instead of generalizing it."

This lack of accordance is attributed by Terman to some fault in method on the part of Swift. "I should judge," he says, "that his results would have been different if he had been able to take his cases individually, instead of collectively."

By way of illustration, the following quantitative results may be quoted: Swift reports that, in the fable mentioned, 27% of his 'bright' group, as contrasted with 9% of his 'dull' group, thought the plan wise and just. Again, only 15% of the bright children, as contrasted with 30% of the dull children, pointed out that the plan would not help the fishes that were not turned into pike. When scored according to Terman's plan[4] his bright *S*'s averaged for three fables the ranks 1.3, 3.17 and 2.83, respectively, while his dull *S*'s averaged the ranks 1.86, 4.57 and 4.57 for the same three fables.

Swift states that "the answers from the public-school children lacked individuality; they were conventional, while those from the reform

[4]In this earlier work Terman gave the score 1 for a satisfactory answer, 5 for a complete failure.

school gave evidence of spontaneity and resourcefulness," and he draws from this the dubious conclusion that "the question may be seriously raised whether the schools do not train children to stupidity."

Quite on the contrary, Terman and Childs cite the following replies to Fable VII given by pupils 13 to 17 years of age, all of whom were retarded in school from 2 to 4 years:

Fable VII.

"The fox was slicker than the crow was."
"Not to be generous to people you don't know."
"Not to sing when you have anything in your mouth."
"To eat before you sing."
"Not to be forgetful."
"Where there's a will there's a way."
"To eat the meat and then sing."
"How to be wise."
"Don't answer if your mouth is full."
"Look before you leap."
"When you have a thing, hang on to it."
"She should not have opened her mouth."
"Teaches us to look for tricks."
"To finish one thing before we do another."
"Taught the crow to be wise and not to open her mouth when she had anything in it."

(4) *Reliability.* No statistical constants have been reported that might serve as coefficients of reliability. It may be mentioned, however, that "an *S*'s previous familiarity with the fables does not necessarily increase in the least his chance of winning a high score." In one room of the Palo Alto schools the 35 pupils had read some or all of the test fables, but these children made no better scores than others of their age and school grade. Terman and Childs conclude that even had an attempt been made to teach the moral of these fables, it would not have been successful if the situation in the fable was naturally beyond the child's powers of comprehension.

(5) *Correlations.* Wyatt reports the following correlations with interpretation of fables: analogies .74, completion test .69, word-building .47, part-wholes test .56, sentence construction .53, memory for nonsense syllables .41, dissected pictures .26, letter-squares .31.

Notes.—For other tests that present a certain analogy to the interpretation of fables, the reader may see Bonser's interpretation of poetry (literary interpretation), already mentioned, Mrs. Squire's tests of supplying a suitable name to, or of ask-

ing appropriate questions about a number of pictures, and Abelson's test of interpretation of pictures.

REFERENCES

(1) A. R. Abelson, Tests for mental deficiency in childhood. *The Child*, 3: 1912, 1-17.

(2) F. G. Bonser, The reasoning ability of children. *ColumbiaCon Ed*, No. 37. New York, 1910.

(3) Carrie R. Squire, Graded mental tests. *JEdPs*, 3: 1912, 363-380, etc., especially 373-376.

(4) E. J. Swift, Standards of efficiency in school and in life. *PdSe*, 10: 1903, 3-22, especially 3-6.

(5) L. M. Terman, Genius and stupidity: a study of some of the intellectual processes of seven 'bright' and seven 'stupid' boys. *PdSe*, 13, 1906, 307-373.

(6) L. M. Terman and H. G. Childs, A tentative revision and extension of the Binet-Simon measuring scale of intelligence. *JEdPs*, 3: 1912, 61-74, 133-143, 198-208, 277-289, especially 133-143.

(7) S. Wyatt, The quantitative investigation of higher mental processes. *BrJPs*, 6: 1913, 109-133.

CHAPTER XII

Tests of Intellectual Equipment

The tests of this chapter differ from other mental tests described in the present volume in that they measure, not the efficiency with which certain typical mental activities or mental processes can function, but rather the number of ideas that an individual possesses. In other words, their purpose is not to measure what the individual can do, or how well he can do it, but what he knows about—to take a census, as it were, of his stock of information. G. Stanley Hall's study of the content of children's minds on entering school[1] is, perhaps, most nearly allied in type and conception with the tests which are here presented.

The first test is designed to secure an estimate of the number of words in the reading vocabulary of the individual tested, the second to secure an estimate of the number of subjects (disciplines, phases of human activity) with which the individual has an exact or an approximate acquaintance.

TEST 50

Size of vocabulary.—Since nearly all thought and expression is couched in linguistic form, and since the intellectual progress of the child at school is, in a sense, a process of augmentation of his vocabulary and of refinement in its use, it seems not unreasonable to assume that the determination of the size of this vocabulary will be of significance and value in estimating his general intellectual status.

Experiments conducted by Kirkpatrick have shown that an approximate determination of what might be termed the *vocabulary-index* can be secured by the use of the relatively short and simple method that is described herewith. By extending the scope of the tests, the usual comparative study may be made, and the index may be related to its conditioning factors—age, sex, school standing, extent of reading, general ability, etc.

[1]See his *Aspects of child life and education.* Boston, 1907.

Kirkpatrick's original list of words has been applied by the author with some modifications of method. Terman and Childs have prepared an entirely different list, on the ground that the words should be selected from a smaller dictionary—one limited to representative and more generally employed words and not including any large proportion of technical terms. They have prepared a list of 100 test-words by taking the last word in every 6th column of Laird and Lee's Vest-Pocket Webster Dictionary, 1904 edition. This dictionary contains 18,000 words, though advertised to contain 30,000, whereas the Webster's Abridged Dictionary used by Kirkpatrick contains 28,000 words.

MATERIAL.—Two printed vocabulary tests: the Kirkpatrick list and the Terman and Childs list (modified by the author in respect to instructions for group testing).

PRELIMINARIES.—In accordance with Kirkpatrick's plan, several preliminary exercises are employed, in order, on the one hand, to obtain data with regard to *S*'s general familiarity with words, his range of reading, etc., and on the other hand, to instill in him an attitude of caution in undertaking the vocabulary-test proper. These preliminary exercises are as follows:[2]

(1) Ask *S* to write the opposite of the following terms: *good, long, break, rude, simple, permanent, particular, permit, obnoxious, genuine.*

(2) Ask *S* to tell (orally or in writing) what the following words mean: *abductor, baron, channel, decemvirate, eschar, amalgamation, bottle-holder, concatenate, disentomb, filiform, gourd, intercede, matting, page, hodman, lanuginose, muff, photograph, scroll, tycoon.* (Where words have more than one meaning, all are to be given.)

(3) Secure from *S* a list of all the papers and magazines that he is in the habit of reading.

(4) Secure from *S* the names of the books that he has read during the past 6 months.

(5) Ask *S* which of these books he liked the best, why he liked it, and to give some account of what it was about.

[2]To follow the plan, the first two exercises, at least, should be given whenever grade pupils are tested, and all five if time permits.

(6) Ascertain the birthplace of *S*'s parents, his school grade, and his favorite school subjects.

METHOD.—Both tests can be conducted by handing the printed forms to *S* and asking him to read the instructions over twice and then to mark the words carefully in accordance with them. But to secure data that will be directly comparable with those published for their test by Terman and Childs the examination of *S* must proceed orally. *S* sees the word, hears it pronounced by *E*, and then gives its meaning orally. *E* scores each term 0, 0.5 or 1. The score 1 is given when *S* is able to give any single correct meaning for the word, even if the meaning given is not the commonest one and even if the definition be poorly expressed. *E* must err on the side of leniency and make due allowance for the difficulty of definition in the case of children.

The following illustrations will indicate the degree of latitude allowed by Terman and Childs in scoring their test: Full credit (1 point) was given for: *Afloat*—"a ship floats on the water;" *civilly*—"it's when you treat a person nice;" *hysterics*—"you act funny or crazy;" *majesty*—"what you say when you are speaking to a king;" *copper*—"something you make money out of." Half credit (0.5 point) was given for: *sportive*—"to like sports:" *pork*—"a kind of meat." It will be seen that a very liberal standard has been used. "Questioning for the sake of drawing out meanings was not resorted to except in rare instances to overcome the child's timidity."

TREATMENT OF DATA.—The Kirkpatrick list supplies *E* directly with the number of words marked 'plus,' and this number indicates the vocabulary-index. For comparison with Kirkpatrick's and Whipple's norms based on this test, the index, taken as a per cent., is multiplied into 28,000.

The Terman and Childs list, when scored directly by *E* upon the oral responses of *S*, gives a vocabulary-index by the simple addition of points scored, and this index, taken as a per cent., is multiplied into 18,000 to make direct comparison with the absolute size of vocabulary as computed by these authors.

The Terman and Childs list, when scored by *E* on a basis of *S*'s own marking (following the author's set of instructions), yields four quantities—number of words that can be defined, that can be explained, that are roughly familiar and that are unknown. The equating of these gradings with the 'points' used by Terman and Childs may be roughly accomplished by

regarding each 'D' and 'E' as indicating one point and each 'F' as indicating a half-point.

VARIATIONS OF METHOD.—To study *S*'s tendency to overestimate or underestimate his vocabulary, *E* may follow the plan used by the author with college students, of giving the check-definition test after the vocabulary-test.

(1) Give *S* the Kirkpatrick vocabulary test without suggesting that he may be called upon to justify his marking.

(2) When the marking has been completed, and the slip is in *E*'s hands, submit to *S* the following list of words, with a request that each word be defined. Allow 20 min. for written definitions.

DEFINITION-LIST.[3]

abductor	interdict	amalgamation	lanugo
abet	interim	amanuensis	lanyard
baroscope	mattock	amaranth	mufti
chanticleer	maturate	bottomry	photo-lithograph
chaos	pudgy	concatenate	rejoinder
decemvirate	scruff	disentrance	skysail
eschar	scrunch	disepalous	tendinous
escheat	subcutaneous	disestablish	tendril
eschalot	tycoon	filiform	virago
gourd	tymbal	hoecake	virescent

(3) For each *S*, ascertain from the definition-test: (*a*) the number of words not defined, (*b*) the number of words wrongly defined. (*c*) Add these to find the total number of words unknown in the list of 40. (*d*) Consult the vocabulary test-slip to see whether any words outside the list of 40 are marked unknown.[4] (*e*) Consult the vocabulary-slip again to see whether any words thereon are marked doubtful and have not been cleared up by the definition-test; consider these as unknown. (*f*) Add all the unknown terms to determine the final corrected vocabulary-index. (*g*) Compare this index with the index indicated by *S* on the vocabulary-slip to see whether *S* has over- or underestimated his vocabulary, and to what degree.

[3] *Concatenation*, *lanuginose* and *lanuginous*, of the vocabulary-test, can, of course, be checked off by the definitions given for *concatenate* and *lanugo*.

[4] In a test of Sophomores and Juniors in college we were surprised to find the following words in this category: *barque*, *barouche*, *boudoir*, *disentomb*, *filigree*, *hodman*, *pagoda*, *rejuvenate*, *scroll*, *sub-let*, *tenderloin*. These words, then, it seems, would have to be added to the 40 to secure a comprehensive list of possibly unknown words.

Results.—(1) Kirkpatrick's computation of the *average vocabulary* is shown in Table 138. The author's results, derived with Kirkpatrick's list applied to 70 college students (16 men and 54 women), aged 16 to 25 years, indicate an average vocabulary of 21,728 when computed on the uncorrected estimates of the students, and of 20,512 when computed on the corrections supplied by the supplementary definition-test.

TABLE 138

Average Vocabulary in Relation to Scholastic Status (Kirkpatrick)

SCHOLASTIC STATUS	VOCABULARY	SCHOLASTIC STATUS	VOCABULARY
Grade II	4480	Grade IX	13,400
Grade III	6620	High school, 1st year	15,640
Grade IV	7020	High school, 2d year	16,020
Grade V	7860	High school, 3d year	17,600
Grade VI	8700	High school, 4th year	18,720
Grade VII	10,660	Normal-school pupils	19,000
Grade VIII	12,000	College students	20,120

The results obtained by Terman and Childs by the use of their list with individual, oral responses from 161 children, aged 5 to 13 years, are shown in Table 139. As will be understood from the explanations already given, the vocabulary-index is larger than Kirkpatrick's, but the absolute vocabulary is smaller.

TABLE 139

Relation of Vocabulary to Age: Method of Terman and Childs

MEDIAN AGE	NUMBER TESTED	MEDIAN INDEX	MEDIAN VOCAB.	VOCAB. REACHED BY 66 PER CENT.	REVISED NORMS	
					Ages	Index
6.5	5	13.9	2500	2300	6	12
7.5	14	14.4	2600	2300	7	14
8.5	28	22.0	3960	3600	8	18
9.5	35	27.8	5000	4000	9	23
10.6	24	33.3	6000	4500	10	26
11.5	29	33.9	6100	5500	11	30
12.4	19	42.9	7700	6500	12	36
13.0	7	48.9	8800	7400	13	42

The author has used the Terman and Childs list with 10 members of the George Junior Republic, ages 14 to 18: the average index was 53.8 per cent. He has also employed the same list by the written response (group method) with Sophomores at Cornell University, with the results indicated in Table 140.

TABLE 140

Vocabularies of Twenty College Students: Terman and Childs List (*Whipple*)

	NUMBER	D	E	F	N	ESTIMATED INDEX
Men	10	65.1	20.8	6.7	7.4	89.3
Women	10	56.3	29.2	8.5	6.0	89.8

On the vocabularies of children below the age of 6 consult Whipple (13) for the chief studies prior to 1908, and Boyd (4), Bush (5), Ghéorgov (6) and Heilig (7) for studies subsequent to that date. For methods and results of securing vocabularies from imbeciles and other feeble-minded consult Binet and Simon (3) and Town (11).

(2) In the author's *definition-test*, no word of the 40 was correctly defined by every student, and since, as has been noted, there remained 16 other words that were unknown or doubtful, it follows that only 44 of the 100 words in Kirkpatrick's list were certainly known by every one of 70 college students.

(3) There is wide *individual variation* in the size of the vocabularies of students of the same age and scholastic status.

TABLE 141

Distribution of Corrected Vocabulary Index. Seventy College Students (*Whipple*)

INDEX	55–59	60–64	65–69	70–74	75–79	80–84	85–89
No. of cases	1	6	13	22	19	5	5

Highest index, 89%. Average index, 73.26%. Lowest index, 58%.

This variation is shown by the distribution in Table 141. The largest college-student vocabulary found by the writer with Kirkpatrick's list is 24,920 (89 per cent.); the smallest is 16,240 (58 per cent.), or approximately the vocabulary assigned by Kirkpatrick to the average 2d-year high-school pupil.

(4) No positive *sex differences* have been established, though there is a suggestion of superiority of boys over girls, and of men over women.

(5) In general, pupils that read the most books and magazines have the largest vocabularies.

(6) Kirkpatrick found a tendency toward positive *correlation* between class standing (teachers' grades) and vocabulary-index: "those ranking high in scholarship knew on an average about 5 per cent. more words than those ranking low in scholarship." The author found a more decided correlation ($r = + 0.45$, P.E. $= 0.06$) between the index of 58 college students and their grades in his classes in educational psychology.

(7) When no precautionary measures are taken to offset the tendency, the determination of the vocabulary-index is commonly affected by *overestimation*. Inspection of Table 142 will show that 59 of the 70 college students examined by the author overestimated, while but 10 underestimated their vocabulary: the largest overestimation was 18 per cent.; the largest underestimation was 4 per cent. Since 20, or more than one-quarter

TABLE 142

Overestimation of the Vocabulary Index. Seventy College Students (Whipple)

PER CENT. OVERESTIMATED	NUMBER	PER CENT. OVERESTIMATED	NUMBER
18	1	5	7
15	2	4	6
14	3	3	7
13	1	2	10
12	3	1	9
11	1	0	1
9	2	—1	6
8	2	—2	2
7	1	—4	2
6	4		

of the students overestimated by 5 per cent. or more, it is evident that, without a somewhat elaborate definition check, the reliability of the vocabulary test is distinctly lessened.[5]

(8) The definition test reveals an unexpectedly large number of *erroneous definitions.* The source of these errors may frequently be traced to confusion with words of similar appearance or to fancied etymological derivations. The following list shows typical errors in definition by college students; the assumed source of confusion is indicated by the terms in parentheses after the definitions:

amanuensis—poet laureate, lovingness (amativeness).
amaranth—a precious stone (amethyst).
abet—although (albeit), a wager (a + bet), diminish (abate).
bottomry—the art of bottoming chairs, deceit, bottom of anything.
chanticleer—one who sings a loud song, one who leads a chant.
decemvirate—composed of five, count out by tens, formerly a group of ten men, but any number now.
disentrance—failure to enter.
disepalous—apart from the head, without shoulders (di + cephalous?).
gourd—reward (guerdon), to slash or whip (goad), morning glory.
interim—time between two reigns (interregnum).
lanugo—a kind of language.
lanyard—yard where leather is tanned (tanyard), yard about the lane.
mattock—a lock of hair (matted locks?), a kind of bird, a sort of rug, a kind of robe (cassock).
maturate—to ripen (mature), to matriculate.
sky-sail—a sail in the sky, a kite.
tycoon—a violent wind (typhoon), an animal, a silk-worm, a natural phenomena (sic).
tendril—a membrane connecting two bones (tendon).
tendinous—capable of endurance (tenacious?).
scrunch—a good for nothing person (scrug?).
virago—a kind of bird (!) (vireo), a disease, giddiness (vertigo).
virescent—sparkling (iridescent), of or pertaining to man (!) (virile).

NOTES.—(1) The greatest source of unreliability in the vocabulary tests in which *S*'s mark their own papers lies in individual differences in the subjective standard employed by different *S*'s

[5]This result may be compared with Kirkpatrick's conclusion that very young children are apt to underestimate because the isolated words of the list fail to arouse associations such as they would if they had a context. Again, when Kirkpatrick defined the words of the list to normal-school students, he found that the errors of over- and underestimation tended to cancel one another; while when college classes defined 20 words, 114 of 246 students (about 46 per cent.) correctly defined the same proportion that they had marked as known, and only 7 per cent. erred by as much as 3 in 20.

by the 'known'-or-'unknown' method: some *S*'s mark, as known, words which are little more than familiar; others mark words as known only when they can define them accurately.

(2) This leads one to say again that, especially in the case of young children, there may be a tendency toward underestimation of the vocabulary because isolated words sometimes fail to arouse the interpretative meanings that they would arouse at once in their customary context. In so far as appeal to the ear as well as to the eye is of assistance to young children whose vocabulary is largely auditory, this source of error is partially offset by the procedure adopted by Terman and Childs of reading the list aloud to them.

(3) In grading the definition test, it is at times rather difficult to decide from the definitions whether *S* does or does not know the meaning of a word with sufficient exactness to be credited with knowledge of the term in question. In general, it is better, in consideration of the difficulty of accurate definition and of the short time usually available for this part of the test, to err on the side of leniency.

Thus the following definitions might be accepted: 'disestablish—to overthrow,' 'decemvirate—a body of ten,' 'mattock—a garden tool,' 'amaranth—a flower;' while the following ought, in our opinion, to be disallowed: 'lanyard—one of the spars of a ship,' 'decemvirate—Roman civil officer,' 'gourd—a hollow vessel from which to eat and drink,' 'concatenate—to argue,' 'baroscope—an instrument for measuring something.'

(4) The pamphlet issued by Ayres (1) is of interest as showing how few words, relatively speaking, are employed in the conduct of ordinary correspondence. Out of a total of 23,629 works, taken by the method of random samples from 2,000 letters, there were only 2,001 different words.

REFERENCES

(1) L. P. Ayres, The spelling vocabularies of personal and business letters. (No. E 126, Russell Sage Foundation.) New York, 1914. Pp. 14.
(2) E. H. Babbitt, A vocabulary test. *PopSciM*, 70: 1907, 378.
(3) A. Binet et Th. Simon, Langage et pensée. *AnPs*, 14: 1908, 284-339.
(4) W. Boyd, The development of a child's vocabulary. *PdSe*, 21: 1914, 95-124.

(5) A. D. Bush, The vocabulary of a three-year-old girl. *PdSe*, 21: 1914, 125-142.
(6) I. Ghéorgov, Le développement du langage chez l'enfant. (Reprint from First Intern. Cong. Pedology.) Ledeberg-Gand, 1912. Pp. 18.
(7) M. R. Heilig, A child's vocabulary. *PdSe*, 20: 1913, 1-16.
(8) E. Kirkpatrick, Number of words in an ordinary vocabulary. *Sci*, 18: 1891, 107-8.
(9) E. Kirkpatrick, A vocabulary test. *PopSciM*, 70: 1907, 157-164.
(10) L. M. Terman and H. G. Childs, A tentative revision and extension of the Binet-Simon measuring scale of intelligence. *JEdPs*, 3: 1912, 61-74, 133-143, 198-208, 277-289, especially 205-208.
(11) Clara H. Town, A study of speech development in two hundred and eighty-five idiots and imbeciles. *JPsAs*, 17: 1912, 7-15.
(12) G. M. Whipple, Vocabulary and word-building tests. *PsR*, 15: 1908, 94-105.
(13) G. M. Whipple and Mrs. Whipple, The vocabulary of a three-year-old boy, with some interpretative comments. *PdSe*, 16: 1909, 1-22. (Contains references to 27 articles on children's vocabularies.)

TEST 51

Range of information.—The words that comprise Kirkpatrick's vocabulary test are intentionally selected by chance: some of them, like *page,* happen to be most ordinary and everyday terms; others, like *lanuginose,* are unusual, technical terms. The extent of *S*'s acquaintance with words of the latter kind depends almost entirely upon the nature of his school training, or upon the quantity and type of his general reading.

The range of information test has been devised by the author as an extension of the vocabulary test. The hundred test-words have been selected, not by chance, but by careful consideration, and in such a manner that each shall be representative of some specific field of knowledge or activity, in the sense that if *S* has made himself familiar with a given field, he will almost certainly know the word selected from that field, whereas if he has not made himself familiar with the field, he will almost certainly not know the term, or at least will not have such knowledge of it as to enable him to define it exactly. Thus, general knowledge of American history is tested by the name *Anthony Wayne,* knowledge of French by *aujourd'hui,* of chemistry by *chlorine,* of ethics by *hedonism,* of golf by *midiron,* of social usages by *R. S. V. P.,* of the technique of photography by *f-6.4,* etc.

MATERIAL.—Specially prepared test-blank containing 100 test-words, directions for marking them, and a request for 10 definitions.

METHOD.—Place the blank in *S*'s hands; ask him to read the directions through twice before marking the words, and call his attention to the request for definitions as printed below the test-words. Let him take his own time.

For exact results, *S* should afterward be required to define every word that he has marked *D*, and to explain or attempt to explain every word that he has marked *E* or *F*. This check test should, by preference, be conducted orally. In practise, however, especially when testing by the group method, such careful checking may prove too onerous; erroneous definitions may then be neglected, or the quantitative data may be revised by discounting on the basis of the percentage of error revealed in the definitions actually given. Or, again, *E* may, after the test is concluded, define the 100 words, and let each *S* revise his own paper by placing a second series of marks *after* each word to indicate the manner in which he should have marked it. A comparison of the sums of the *D*'s, *E*'s, *F*'s and *N*'s of the first and of the second series will then show approximately the extent and nature of the error due to ignorance or misunderstanding of the real meanings.

RESULTS.—(1) *Dependence on school training.* Results obtained by the author at Cornell University and the Ithaca, N. Y., High School, and by Miss Smith at the University of Texas are presented in Tables 143 and 144. In both it is evident that advance in school training, together, of course, with in-

TABLE 143

Dependence of Range of Information on Academic Status (Whipple)

ACADEMIC STATUS	NUMBER	D	E	F	N
Graduates	4	39.0	21.0	12.2	27.0
Seniors	5	20.6	17.2	25.2	37.0
Juniors	10	24.8	12.0	23.7	39.5
Sophomores	30	17.7	12.7	17.3	52.2
High School	52	6.8	7.6	16.3	69.3

creased maturity, is paralleled by an increase in the number of technical terms that can be defined (*D*), explained (*E*), or that are at least familiar (*F*), and by, naturally, a corresponding decrease in the number of terms that are new or unknown (*N*).

TABLE 144

Dependence of Range of Information on Academic Status (Smith)

ACADEMIC STATUS	NUMBER	D	E	F	N
Graduates	9	38.11	12.11	15.67	34.11
Seniors	47	24.90	16.70	19.30	39.10
Juniors	59	20.50	14.40	20.70	44.40
Sophomores	85	20.10	12.10	19.10	48.70
Freshmen	153	13.70	10.70	15.50	60.10

(2) *Dependence on sex.* The results obtained at Ithaca and Austin, reclassified by sex, are shown in Table 145, where it is evident that there exists a superiority of range of information in the males.

TABLE 145

Dependence of Range of Information on Sex (Whipple and Smith)

	NUMBER	D	E	F	N
Ithaca men	44	15.79	11.98	18.22	54.02
Ithaca women	57	12.21	9.42	17.19	61.17
Texas men	162	21.00	12.70	15.60	50.70
Texas women	173	15.20	13.20	19.30	52.30

(3) The results just figured are 'raw' results; strictly speaking, these should be revised on the basis of an extended series of definitions, as recommended in the Vocabulary Test (No. 50), since an inspection of the definitions and explanations actually given reveals in the majority of the papers one or more errors,

due in the main to confusion with words of similar appearance or to fancied etymological derivations. The following list shows typical errors in definition by college and high-school students; the assumed source of confusion is indicated by the terms in parenthesis after the definitions:

ageratum—an aggregation of objects: the aggragate (sic) amount.

annealed—pressed or rolled out thin: molded together.

Anthony Wayne—a historic character who was hung in the cause of freedom for the blacks: a man who fought in the Revolution on the English side.

Babcock test—a device to ascertain whether or not cattle have tuberculosis.

base-hit—when the ball is hit and strikes a base or is caught there: a ball batted over a base: when the striker bats the ball into the pitcher's hands.

Bokhara—name of a place in Austria.

cantilever—a bar with a hook in one end by which lumbermen roll logs (canthook).

catalepsy—a form of disorder of the nervous system which causes fits or convulsions (epilepsy). (Similar statements given by 15 persons.)

chamfer—the tree from which camphor gum is obtained: this is the simplified spelling of it (!). (The confusion with camphor was found in 4 papers.)

clearing-house—a sale that takes place when a store wishes to dispose of its stock (clearing sale): a place where clearing papers are given to vessels to enable them to leave the harbor (customs house + clearing of vessels): picking up everything to move; taking everything out of a house: a place used by express companies to sell uncalled-for goods: a house where goods are made ready to be delivered.

cotangent—name of one of two tangents drawn to a circle from the same point without the circle: one lying alongside of (contingent): straight line drawn to touch a circle at one point (tangent).

dibble—to get just a smattering of some subject, as to dibble in medicine or politics (dabble): to do with divided interest (dawdle).

dryad—a priest of early English times (druid).

entrée—first course at a banquet, usually soup: something in the way of food, new and out of season: when the waiter brings in a new course it is called an entré: French for 'to-day': French for 'between' (*entre*).

Eocene—the term applied to one of the early ages of civilization.

Euclid—a book written by Vergil (Æneid): name given to certain trees (eucalyptus): an ancient Egyptian who studied geometry: name of an avenue in Cleveland, Ohio.

f-64—means the temperature is 64 degrees above zero, Fahrenheit.

f. o. b.—cash on delivery (c. o. d): forward on board.

golden section—the section of the West most prosperous.

hydraulic press—a kind of air-pump, rather complicated, operated by suction and pressure: a machine for washing dirt from gold or from steep slopes (hydraulic mining): the force with which water flows upon or against a thing, as a paddle wheel.

impressionism—when a man imitates the looks or actions of another: the art of exciting an impression.

infusoria—a chemical herb (infusion?).

kilogram—the greatest quantity in the metric system: French measure of distance (kilometer): French unit of liquid measure: the weight of a cube of water whose dimensions are a kilometer.
Les Misérables—a French tragedy written about the last part of the 17th century by Racine, one of the famous French writers: French work written by George Sand, author of Le Diable.
linotype—the product of a certain method of making prints from photographs.
Millet—a blind poet (Milton).
natural selection—in nature each animal selects its mate, a device for building up a stronger race.
ohm—German word for uncle (*Oheim*).
Polonius—a prominent character in Julius Caesar.
pomology—the study of the palm of the hand, used by fortune tellers (palmistry).
tort—French word for ugly (*tors?*).
triple expansion—the expanding of anything three times its normal size.
Utopia—a silk factory.
way-bill—a bill that is being considered.
Zionism—same as Dowieism.

(4) A comparison of scores made by 18 summer-session students, before and after the definition by E of the 100 terms, shows the following averages: first marking, $D = 20.39$, $E = 14.77$, $F = 18.39$, $N = 46.44$; second marking, $D = 19.77$, $E = 20.22$, $F = 19.55$, $N = 40.44$. So far as these S's are concerned, then, it appears that at first they had overestimated terms definable and, more particularly, terms unknown. The principal effect of E's explanations was to increase by about 6 per cent. the number of terms marked as explainable, and to decrease by 6 per cent. the number of terms marked as unknown.

NOTE.—Attention may be called to the suggestive method devised by Franken (1), the purpose of which is to test not so much the range of information of pupils, but rather the degree to which they overestimate their range and the extent to which this overestimation may be reduced by proper drill and instruction. A series of questions drawn from school work is propounded, first in a form that inquires as to the existence of the information and that requires merely the answer 'yes' or 'no.' After this series has been answered, the same questions are given in a second form that demands a specific answer. Examples: first form: "Do you know what city is the capital of France"? Second form: "What city is the capital of France"? For various ways of conducting tests by these two forms of

questions the original articles should be consulted. Franken's method is in many features more akin to the 'Aussage' test (No. 32).

REFERENCES

(1) A. Franken, (*a*) Ueber die Erziehbarkeit der Erinnerungsaussage bei Schulkindern. *ZPdPs*. 12: 1911. 635-642. (*b*) Aussageversuche nach der Methode der Entscheidungs- und Bestimmungsfrage bei Erwachsenen und Kindern. *ZAngPs*, 6: 1912, 174-253.

(2) Laura L. Smith, Whipple's range of information test. *PsR*, 20: 1913, 517-518.

(3) G. M. Whipple, A range of information test. *PsR*, 16: 1909, 347-351.

CHAPTER XIII

Serial Graded Tests for Developmental Diagnosis

The omission from this volume of the de Sanctis tests and of the Binet-Simon tests demands a brief explanation. The reasons that have led to this omission are: first, the extension of the material of the preceding pages has brought the volume to dimensions already in excess of the original plans; secondly, the number of published investigations bearing upon the Binet tests is so enormous (Kohs' bibliography lists 254 titles to June, 1914) that the proper consideration of so much material demands more time than can be permitted; thirdly, the extensive use of the Binet tests has given rise to so many variations in method of application and scoring that there now exist numerous issues upon each one of which an authoritative presentation must take a definite and justified stand—something which is impossible without extensive comparison of the views of various writers and resolution of the conflicting views on the basis of careful first-hand investigation; fourthly, there are now available a number of pamphlets of directions prepared by competent writers (Goddard, Kuhlmann, Schwegler, Terman, Town, Winch, *et al*), so that the need for a Binet handbook that was felt when the first edition of this *Manual* appeared is now sufficiently met, and it would be only adding confusion to present still another version of the tests if it were set forth without sufficient justification to claim attention as a standardized version.

It is my hope, however, to issue later a supplementary volume that will discuss the rationale of combinations of tests into systems, that will deal with the Binet tests in a comprehensive manner, and that will include also other systems of tests, such as the de Sanctis tests and the psychological-profile method of Rossolimo. In the meantime, the selected references that follow will serve to guide the reader to some of the more important discussions in English of two of these test-systems.

REFERENCES

(1) J. C. Bell, Recent literature on the Binet tests. *JEdPs*, 3: 1912, 101-110.

(2) C. S. Berry, A comparison of the Binet tests of 1908 and 1911. *JEdPs*, 3: 1912, 444-451.

(3) H. H. Goddard, (*a*) The Binet-Simon measuring scale for intelligence. Revised. *TrSc*, 8: 1911, 56-62. (*b*) Two thousand normal children measured by the Binet measuring scale of intelligence. *PdSc*, 18: 1911, 232-259. (*c*) Standard method for giving the Binet test. *TrSc*, 10: 1913, 23-30.

(4) E. B. Huey, (*a*) The Binet Scale for measuring intelligence and retardation. *JEdPs*, 1: 1910, 435-444. (*b*) The present status of the Binet scale of tests for the measurement of intelligence. *PsBu*, 9: 1912, 160-168. (*c*) Backward and feeble-minded children. *EdPsMon*, 1912. Pp. 221.

(5) S. C. Kohs, The Binet-Simon measuring scale for intelligence: an annotated bibliography. *JEdPs*, 5: 1914, 215-224, 279-290, 335-346. (Also sold separately.)

(6) F. Kuhlmann, (*a*) A revision of the Binet-Simon system for measuring the intelligence of children. *Mon. Suppl. of JPsAsth*, 1: 1912. Pp. 41. (*b*) Some results of examining a thousand public school children with a revision of the Binet-Simon tests of intelligence by untrained examiners. *JPsAsth*, 18: 1914, 150-179, 233-269. (Also published separately.)

(7) S. de Sanctis, Mental development and the measure of the level of intelligence. *JEdPs*, 2: 1911, 498-507.

(8) R. A. Schwegler, A teachers' manual for the use of the Binet-Simon scale of intelligence. (Selected bibliography of 56 titles.) *Univ. of Kansas, School of Education*, 1914. Pp. 56.

(9) W. Stern, The psychological methods of testing intelligence. *EdPsMon*, No. 13, 1914. Pp. 160.

(10) L. M. Terman, Suggestions for revising, extending and supplementing the Binet intelligence tests. *Intern. Conf. School Hyg.*, Buffalo, N. Y., 1913.

(11) L. M. Terman and H. G. Childs, A tentative revision and extension of the Binet-Simon scale of intelligence. *JEdPs*, 3: 1912, 61-74, 133-143, 198-208, 277-289.

(12) Clara H. Town (translator), A method of measuring the development of the intelligence of young children, by A. Binet and Th. Simon. Lincoln, Ill., 1913.

(13) J. E. W. Wallin, (*a*) Experimental studies of mental defectives. A critique of the Binet-Simon tests. *EdPsMon*, No. 7, 1912. Pp. 155. (*b*) The mental health of the school child. New Haven, 1914. Pp. 463.

APPENDIX I

Formulas for Converting Measures (English and Metric Systems)

Measures of Length

1 mm. = 0.0394 inch.
1 cm. = 0.3937 inch.
1 m. = 39.37 inches.
1 in. = 2.54 cm.
1 ft. = 0.3048 m.

Measures of Surface

1 sq. cm. = 0.155 sq. in.
1 sq. in. = 6.452 sq. cm.

Measures of Capacity

1 cu. cm. = 0.061 cu. in.
1 cu. in. = 16.4 cu. cm.

Measures of Weight

1 gram = 0.035 oz.
1 kg. = 2.204 lbs.
1 oz. = 28.35 g.
1 lb. = 453.59 g.

APPENDIX II

List of Abbreviations

The following abbreviations, save for a few additions, are identical with those recommended and employed in the *Zeitschrift für angewandte Psychologie*, V, Heft 5-6, VI, Heft 5-6.

AmAnt: American Anthropologist (Lancaster, Pa.).
AmJIns: American Journal of Insanity (Baltimore, Md.).
AmJPhg: American Journal of Physiology (Boston, Mass.).
AmJPs: American Journal of Psychology (Worcester, Mass.).
AmJSci: American Journal of Science (New Haven, Conn.).
AnPs: L'Année psychologique (Paris).
ArGsPhg: Archiv für die gesamte Physiologie des Menschen und der Tiere (Bonn).
ArGsPs: Archiv für die gesamte Psychologie (Leipzig).
ArPs(e): Archives of Psychology (New York).
ArPs(f): Archives de Psychologie (Geneva, Switzerland).
BuAcRoySci: Bulletins de l'Académie Royale des Sciences, des Lettres et des Beaux-arts de Belgique (Brussels).
BerlinKlW: Berliner Klinische Wochenschrift (Berlin)
BiZb: Biologisches Zentralblatt (Erlangen).
BrJPs: British Journal of Psychology (Cambridge, England).
BuSocEtPsEnf: Bulletin de la Société libre pour létude psychologique de l'enfant (Paris).
ColumbiaConEd: Columbia Contributions to Education (New York).
ColumbiaConPhPs: Columbia Contributions to Philosophy and Psychology (New York).
DMdW: Deutsche Medizinische Wochenschrift (Leipzig).
Ed: Education (Boston, Mass.).
EPd: Die experimentelle Pädagogik (Leipzig).
EdPsMon: Educational Psychology Monographs (Baltimore, Md.).
FsPs: Fortschritte der Psychologie und ihre Anwendungen (Berlin).
InMagScHyg: International Magazine of School Hygiene (Leipzig).

JAntInst: Journal of the Anthropological Institute of Great Britain and Ireland (London).
JEdPs: The Journal of Educational Psychology (Baltimore, Md.).
JEPd: Journal of Experimental Pedagogy and Training College Record (London).
JNeMeDis: Journal of Nervous and Mental Disease (New York).
JPh: Journal of Philosophy, Psychology and Scientific Methods (New York).
JPhy: Journal of Physiology (Cambridge, England).
JPsAsth: Journal of Psycho-Asthenics (Faribault, Minn.).
NeMeDisMon: Nervous and Mental Disease Monograph Series (New York).
PdPsArb: Padagogisch-psychologische Arbeiten (Leipzig).
PdSe: Pedagogical Seminary (Worcester, Mass.).
PdJb: Paedologisch Jaarboek (Antwerp).
PhR: Philosophical Review (Lancaster, Pa.).
PhSd: Philosophische Studien (Leipzig).
PopSciM: Popular Science Monthly (Garrison, N. Y.).
PsArb: Psychologische Arbeiten (Leipzig).
PsBu: Psychological Bulletin (Lancaster, Pa.).
PsCl: Psychological Clinic (Philadelphia, Pa.).
PsMon: Psychological Monographs (Lancaster, Pa.).
PsR: Psychological Review (Lancaster, Pa.).
RepComEd: Report United States Commissioner of Education (Washington, D. C.).
RMdSuisse: Revué médicale de la Suisse Romande (Geneva, Switzerland).
RPhF: Revue philosophique de la France et de l'Etranger (Paris).
RSci: Revue scientifique (Paris).
Sci: Science (Garrison, N. Y.)
SdYalePsLab: Studies from the Yale Psychological Laboratory.
SmAbPdPs: Sammlung von Abhandlungen aus dem Gebiete der pädagogischen Psychologie und Physiologie (Berlin).
TrSc: The Training School (Vineland, N. J.).
UnIowaSdPs: University of Iowa Studies in Psychology (Iowa City, Iowa).
ZAngPs: Zeitschrift für angewandte Psychologie und psychologische Sammelforschung (Leipzig).
ZBi: Zeitschrift für Biologie (Munich).
ZEPd: Zeitschrift für experimentelle Pädagogik (Leipzig).
ZPdPs: Zeitschrift für pädagogische Psychologie und experimentelle Pädagogik (Leipzig).
ZPs: Zeitschrift für Psychologie (Leipzig).
ZScGd: Zeitschrift für Schulgesundheitspflege (Hamburg).

APPENDIX III

List of Materials

Numerals refer to test-numbers. Items starred refer to materials that are recommended, but not prescribed, or to materials for the conduct of alternative or supplementary tests.

The Materials may be ordered of C. H. Stoelting Company, 3047 Carroll Ave., Chicago, Illinois, who will quote prices on application.

I. SPECIAL APPLIANCES

Card of objects, Binet's, 32
Counter, mechanical, 36*
Demoor suggestion blocks, 40*
Kymograph drum and stand, 42, 43
Memory apparatus, Jastrow's, 38*
Pendulum, seconds', 38*
Pictorial completion test, Healy's, 48*
Pictures (lithographs), Hindoos, 31; Australians, 32; Disputed Case, 32; Washington and Sally, 32*; Orphan's Prayer, 32*; card of 13 colored, 38
Prism, 20-D., 36*
Seconds clock, 33*, 34*, 35*, 48*
Suggestion blocks, set of 22, 40
Warmth illusory, electrical apparatus for, 44; Guidi's stove, 44
Weights, progressive, set of 15 for suggestion, 41

II. SPECIAL PRINTED FORMS

Analogies, 3 sets of stimulus cards and recording blank, 34A; three forms for group tests of same, 34A
Association, 100-word list, 33; see Analogies, Controlled association and Kent-Rosanoff test
Completion test, Ebbinghaus', set of 4 forms for, 48
Computation tests, addition book, 35; addition problems, Schulze's method, 35; addition problems, 2-place digits, 35; addition problems, 20-place digits, 35; multiplication problems, 35
Controlled association, 4 sets of stimulus cards, 20 each, for part-whole, genus-species and opposites (2 forms) tests, 34; four forms for group tests of same, 34
Fables, set of 8, 49
Information test, 51
Ink-blots, set of 20, 45
Kent-Rosanoff association test, 33A; frequency tables for same, 33A
Memory for ideas, 3 test sheets (Marble Statue, Cicero and Dutch Homestead), 39
Memory for letter-squares, set of 10 test-cards and blanks for records, 38
Memory for sentences, 2 test sheets of 21 sentences each, 38
Memory-span for digits, set of 42 test-cards, 38
Mirror-drawing, 6-pointed star, 36; set of 6 patterns, 36
Sentences for completion, 46
Substitution test-strips (Form A), 37; coverboard with key (Form A), 37; test blanks (Form B), 37; set of 4 test-sheets and cardboard key (Form C), 37
Vocabulary tests, 2 forms, 50
Word-building, 2 forms for, 47

III. GENERAL APPLIANCES AND MATERIALS

Alcohol, denatured, 44
Cardboard, 36, 37, 42, 43, 46
Cigarette, 31
Cloth, soft black, 40, 41
Cross-section paper, ruled in mm., 42
Drawing materials, 42, 43
Gummed letters and figures, Sizes 5 and 10, 38*
Lamp, alcohol, 44
Lamps, four 25-watt tungsten, 44
Matches, 44
Metronome, 38
Mirror, 36
Stamp, 2-cent, 31
Stop-watch, preferably split-second, 32-39, 44-48
Supports, 36
Thumb-tacks, 36

INDEX OF NAMES

Roman numerals refer to test-numbers, italicized numerals to page-numbers.

Aall, A., 39
Abbott, Edwina E., 38
Abelson, A. R., 34, 38, 49
Abt, G., 36
Aikins, H. A., 34, 35
Allen, F. J., 36
Anderson, E. J., 38, 47
Angell, J. R., 38
Arai, Tsuru, 35
Aschaffenburg, *518*
Ayers, L. F., 50

Baade, W., 32, 39
Babbitt, E. H., 50
Baginsky, A., 32
Balaban, A., 39
Baldwin, B. T., 37
Bell, J. C., *589f*, 43, *690*
Bellei, 35
Bentley, M., 38
Bergström, J. A., 38
Bernstein, A., 38
Berry, C. S., *690*
van Biervliet, J., 38, 40
Bigham, J., 38
Binet, A., 31, 32, 33, 38, 39, 41, 42, 43, 45, 46, 48, 50
Bingham W. V., *458*
Bischoff, E., 35
Bogdanoff, T., 38
Bolton, F. E., 40
Bolton, T. L., 38
Bonser, F. G., 34, 49
Boring, E. G., 32
Borst, Marie, 32
Bourdon, B., 38
Bovet, P., 33A
Boyd, W., 50
Brand, J. E., 43
Breukink, H., 32
Briggs, T. H., 34, 34A
Brown, W., 35, 38, 48
Brugmans, H., 38, 39, 47, 48
Bryant, Sophie, 31, 33A
Burgerstein, L., 35
Burnham, W. H., 38
Burris, W. P., 35
Burt, C., 33, 33A, 34, 34A, 35, 36, 38, 46, 48
Bush, A. D., 50

Calfee, Marguerite, 36
Calkins, Mary W., 33, 38
Carpenter, D. F., 34, 37, 38
Cattell, J. M., 33, 33A
Chambers, W. G., 34, 35, 38
Charpentiér, A., 40
Childs, H. G., 48, 49, 50, *690*
Chojecki, A., 42, 44
Claparède, E., 32, 40
Clavière, J., 40
Cohn, J., 31, 32, 38, 48
Courtis, S. A., 35

Dallenbach, K., 32
Daniel, 40
Dauber, J., 32
Dearborn, G. V. N., 45
Dearborn, W. F., 36, 37
Dell, J. A., 39
Demoor, 40
Descoeudres, Alice, 31
Dieffenbacher, J., 31, 32, 38, 48
Doll, E. A., 40
Donovan, M. E., 35
Downey, June E., 36
Dresslar, F. B., 40
Duprée, E., 32

Eastman, F. C., 33A
Ebert, E., 38
Ebbinghaus, E., 35, 38, *619*, *640*, 48
Elsenhans, T., 48
Ephrussi, P., 38

Finzi, J., 38
Fischer, Charlotte, 34, 37, 46
Flournoy, Th., 33, 40
Fourche, J. A., 40
Fox, W. S., 35
Franken, A., 32, 51
Fränkl, E., 38
Fraser, D. K., 48
Friedrich, J., 35
Fuchs, H., 38

Galton, F., 38
Gamble, Eleanor, 38
Gerland, H. B., 32
Ghéorgov, I., 50
Gilbert, J. A., 40
Giroud, A., 41, 42
Goddard, H. H., *689f*
Goett, T., 33A
Gray, C. T., 37
Gross, H., 32
Guidi, G., 44

Haggenmüller, A., 38
Hall, G. S., *674*
Hawkins, C. J., 38
Healy, W., 48
Heck, W. H., 35
Hegge, T., 32
Heilig, M. R., 50
Heindl, R., 32
Henderson, E. N., 39
Henmon, V., 38
Henri, V., 31, 36, 38, 39, 43, 46, 48
Heymans, G., 38, 39, 47, 48
Hill, D. S., 36
Hollingworth, H. L., 34, 35, 40
Holmes, Marion, 35
Hubbell, Elizabeth, 35
Huey, E. B., *690*

Itschner, H., 38

Jacobs, J., 38
Jaffa, S., 32
Jastrow, J., 33, *500*
Johnson, G. E., 38
Jones, E., 34, 35
Jones, Grace M., 43
Jones, W. F., 38
Jost, A., 38
Judd, C. H., 36
Jung, C. G., 33A

Kakise, H., 33A
Katzen-Ellenbogen, E., 35
Keller, R., 35
Kelley, T. L., 33A
Kemsies, F., 35, 38
Kent, Grace, 33A
Kirkpatrick, E. A., 38, 45, 50
Kline, L. W., 37
Kohnky, Emma, 37, 38
Kohs, S. C., *689f*
Kraepelin, E., *409*, 35, 38, 48
Krueger, F., 35, 38, 48
Kuhlmann, F., 38, *689f*

Lapie, P., 38
Laprade, A., 36
Larguier des Bancels, J., 38
Laser, H., 35
Lay, W. A., 38
LeClere, A., 31
Levy-Sühl, M., 33A
Ley, 33A
Lipmann, O., 32, 48
Lobsien, M., 32, 38, 48
Lochte, 36
Loomis, H. N., 40
Lough, J. E., 37

Manchester, Genevieve, 33
Martin, Gladys, 35
Martin, Lillien, 40
Masselon, R., 46
Maurer, L., 32
McDougall, W., 38
Menzerath, 33A
Meumann, E., 33A, 34, 35, 38, 46
Meyer, M., 48
Michotte, A., 39
Monroe, W., 31
Moore, R. C., 33, 33A, 34, 34A, 35, 36, 38, 48
Müller, G. E., 38, 40
Munn, Abbie F., 37
Münsterberg, H., 32, 38

Netolitzky, A., 35
Netschajeff, A., 38
Nevers, Cordelia, 33
Norsworthy, Naomi, 34, 38

Oehrn, A., 35
Offner, M., 35, 38
Ogden, R. M., 39
Okabe, T., *602*, *607*, *617*
Ordahl, Louise, 35, 36
Otis, Margaret, 33A

Pearce, H. J., *589f*
Peterson, H. A., 39

Phillipe, J., 40
Pilzecker, A., 38
de Placzek, 32
Pohlmann, A., 38
Portych, T., 39
Powelson, Inez, *590*
Pyle, W. H., 33, 34, 37, 38, 39, 45, 47

Quantz, J. O., 38

Ransy, C., 39
Reinhold, F., 33A
Reis, J., 35
Reuther, F., 38
Rice, J. F., 40
Ries, G., 34
Ritter, C., 38
Robinson, L. A., 35
Roels, F., 33A
Rosanoff, A. J., 33A
Rosanoff, Isabel, 33A
Rowe, E. C., 36
Rusk, R. R., 33A, 34

Saling, Gertrud, 33A
de Sanctis, S., *690*
Schramm, F., 39
Schulze, R., 35
Schultz, G., 32
Schumann, F., 38, 40
Schuyten, M. C., 38
Schwegler, R. A., *689f*
Scott, W. D., *590*, 44
Scripture, E. W., 40, 44
Seashore, C. E., 40, 44
Segal, J., 38
Sharp, Stella, 31, 38, 39, 45, 46
Shaw, J. C., 39
Simon, T., 50
Simpson, B. R., 34, 35, 38, 39, 48
Sleight, W. G., 38
Small, M. H., 44
Smedley, F., 38
Smith, Laura L., 51
Smith, W. G., 38
Sommer, *420*
Spearman, C., 35, 38, 48
Specht, W., 35
Squire, Carrie R., 31, 34, 37, 38, 46, 47, 48, 49
Starch, D., 35, 36, 37, 38
Stern, Clara, 32
Stern, W., 32, *590*, *690*
Strack, M., 36
Stratton, G. M., 36
Strong, E. K., Jr., 33A, *589f*
Swift, E. J., 49
Sybel, A. v., 38

Tanner, Amy, 33
Teljatnik, 35
Terman, L. W., 39, 47, 48, 49, 50, *689f*
Thorndike, E. L., 34, 35
Titchener, E. B., 38
Town, Clara H., 50, *689f*
Toulouse, *520*

Vogt, R., 35
Vos, H. B., 39

Wallin, J. E. W., 33, 39, *690*
Washburn, Margaret F., *590*
Watt, H. J., 34
Weber, H., 36
Wegener, H., 36
Weidensall, Jean, 34, 36, 37, 46
Wells, F. L., 33A, 34, 34A, 35, 37
Wertheimer, M., 48
Wessely, R., 38
Whipple, G. M., 32, *602*, *607*, *617*, 47, *663*, 50, 51
Whipple, Mrs. G. M., 50
Wiersma, E., 48
Wigmore, J. H., 32
Winch, W. H., 35, 38, 39, 46, *689*
Winteler, J., 33A, 34
Wissler, C., 38, 39
Wolfe, H. K., 40
Woodworth, R. S., 33A, 34, 34A, 35, 37
Woolley, Helen T., 34, 37, 46
Wreschner, A., 32, 33A
Wyatt, S., 34, 34A, 35, 38, 46, 47, 48, 49

Yoakum, C. S., 36, 38

Ziehen, T., 33A

INDEX OF SUBJECTS

For authors quoted, see Index of Names; for apparatus, see List of Materials; for tables and figures, see indexes following Table of Contents.

Abnormal children, see Delinquents and Feeble-minded
Accuracy, relation of, to speed, see Speed
Addition, see Computation
Age, effect of, on description, 378 f.; on report, 399 f.; on uncontrolled association, 413; on Kent-Rosanoff test, 426-430; on part-whole test, 442; on genus-species test, 444; on opposites test, 450; on analogies test, 458 f.; on substitution, 506; on immediate memory, 541-545; on logical memory, 578 f.; on size-weight illusion, 593 f.; on progressive-weight test, 601; on progressive line test, 606 f.; on directive suggestion, 611 f.; on warmth illusion, 617; on ink-blot test, 624; on sentence completion, 634; on invention of stories, 636 f.; on word-building, 645; on interpretation of fables, 669 f.
Analogies, 455-460, 662
Apprehension, range of, 383
Arithmetical tests, see Computation and Courtis tests
Association, tests of, 409-485; uncontrolled (continuous method), 410-419; uncontrolled (discrete method), 419-437; controlled (logical relations), 437-455; analogies, 455-460; controlled (computation), 460-485; see Ink-blots
Aussage test, see Report
Ayres' spelling vocabulary, 682

Backward alphabet test, 454
Backward children, see Feeble-minded and General intelligence
Binet-Simon tests, 381, 571 f., 584, 689 f.
Caffein, effect of, on computation, 474
Card of objects, Binet's, 388-393
Cicero test, see Logical memory
Class standing, see General intelligence
Completion test, 627, 640, 649-666
Computation tests, 460-485
Constant-increment test, 465-468
Contradictory-suggestion test, 608 ff.
Correlations, with part-whole test, 442; with opposites test, 453; with computation test, 472 f.; with mirror-drawing, 496 f.; with immediate memory, 565 f.; with logical memory, 584; with progressive-weight test, 602; with progressive-line test, 607; with warmth-illusion, 618; with invention of stories, 637; with development of themes, 639; with word-building, 648; with completion test, 663 f.; with interpretation of fables, 672
Courtis tests, 462, 481
Criminals, see Delinquents

Defective children, see Delinquents, Feeble-minded
Definition test, see Vocabulary test and Information test
Delinquents, performance of, in opposites test, 451; in mirror-drawing, 494 ff.; in substitution, 514 f.; in sentence completion, 634; in completion test, 662 f.
Dental treatment, effect of, on substitution, 514; on memory for digits, 553

De Sanctis' tests, 689 f.
Description test, 376-382
Digit-symbol test, see Substitution
Directive suggestion, 610 ff.
Distraction, effect of, on computation, 475; on immediate memory, 559
Division, see Computation
Dutch Homestead test, see Logical memory

Ebbinghaus test, see Completion test
Environment, influence of, on lists of associations, 418
Epileptics, performance of, in uncontrolled association, 413 f.; in letter-span test, 564; in logical memory, 583 f.

Fables, see Interpretation of fables
Fatigue, effect of, on opposites test, 452; on computation, 475-482; on immediate memory, 552; on logical memory, 581 f.; on completion test, 659 f.
Feeble-minded, performance of, in association test, 429 f., 433 f.; in part-whole test, 442; in genus-species test, 444; in opposites test, 451; in immediate memory, 196 ff.,. with size-weight illusion, 594 f. See Epileptics
Franken's question-method, 687 f.

General intelligence, relation of, to report, 400; to Kent-Rosanoff test, 431 ff.; to part-whole test, 442; to opposites test, 450 f.; to analogies test, 459; to computation, 471; to mirror-drawing, 494; to substitution, 512 ff.; to immediate memory, 560 ff.; to logical memory, 582 f.; to size-weight illusion, 596; to completion test, 661 f.; to interpretation of fables, 670 ff.; to vocabulary, 680
Genus-species test, 442 ff.
Guidi's method for suggesting warmth, 614 f.

Ideational type, effect of, on report, 403; on immediate memory, 557 f.
Illusion, see Suggestion
Imagination, tests of, 619-673
Immediate memory, see Memory
Individual differences, in description, 379 f.; in giving analogies, 459 f.; in computation, 469; in work-curve, 476 f.; in mirror-drawing, 490 f.; in directive suggestion, 611; in ink-blot test, 625; in word-building, 646; in completion test, 659; in vocabulary, 679 f.
Information test, 683-688
Ink-blots, 620-626, 630
Insane, fidelity of report in, 400; free association in, 434; computation in, 474
Instructions, effect of, on Kent-Rosanoff test, 434 ff.
Intelligence, see General Intelligence
Interpretation, of fables, 666-673; of poetry, 672; of pictures, 673
Invention, see Imagination and Stories, invention of

Kent-Rosanoff method, see Association uncontrolled (discrete method)

Learning, tests of, 409 f., 485-516
Letter squares, method of, 528 ff.
Linguistic invention, 626-649
Lipmann-Wertheimer test, 664 f.
Logical memory, 571-587

Marble statue test, see Logical memory
Masselon method, see Sentence formation
Material-weight illusion, 597
Memory, tests of, 383, 409 f., 516-587; for serial impressions (rote memory), 516-570; methods and material classified, 517-521; for digits, 521-528; for letter squares, 528 ff.; for words, 530-534, 554 f.; for sentences, 535 f., 586; for pictures of objects, 536 f.; for ideas, see Logical memory
Mental ability, see General intelligence
Mental work, see Work curve
Meumann's test, see Memory for words and Sentence formation
Mirror-drawing, 485-499
Mirror-writing, 497 f.
Missing-digit test, 482

Multiplication, see Computation
Musical ability, 639

Negro, see Racial differences
Norms, for Kent-Rosanoff categories, 425 f.; for free association times, 425 f.; for part-whole test, 441; for genus-species test, 443 f.; for opposites test, 448 f.; for analogies, 457 f.; for computation, 468 f.; for mirror-drawing, 489 f.; for substitution, 504 ff.; for immediate memory, 537-541; for logical memory, 576 f.; for size-weight illusion, 593 f.; for progressive-weight test, 600 f.; for progressive-line test, 605 f.; for illusion of warmth, 617; for word-building, 644 f.; for completion test, 656 f.; for interpretation of fables, 669; for vocabulary, 678 f.

Opposites test, 445-453

Paralytics, fidelity of report in, 400; see Insane
Part-whole test, 438-442
Perseveration, 427, 435, 533, 567
Physical condition, effect of, on immediate memory, 553
Pictorial completion test, 665
Picture tests, 393-397; see Description, Report, Memory
Pictures, test of titles for, 381 f.
Practise, effect of, on report, 405 f.; on free associations, 431; on opposites test, 452; on computation, 470 f., 482; on mirror-drawing, 491-494; on substitution, 508-512; on immediate memory, 550 ff.; on logical memory, 580 f.; on size-weight illusion, 596; on progressive-weight test, 601; on completion test, 659
Precocity, see General intelligence
Profile method of Rossolimo, 689
Progressive-line test, 602-608
Progressive-weight test, 598-601

Racial differences, in controlled association, 452; in substitution, 506 ff.; in immediate memory, 566; in ink-blot test, 625; in word-building, 645
Range of information, see Information test
Reliability, coefficient of, in opposites test, 452; in analogies test, 459; in computation test, 471 f.; in mirror-drawing, 496; in immediate memory, 559; in logical memory, 582; in invention of stories, 637; in word-building, 647 f.; in completion test, 663; in interpretation of fables, 672
Report, test of, 383-408
Resistance-wire method, for suggesting warmth, 613 f.
Rest-pauses, see Work curve
Ries' association test, 453 f.

School grade, relation of, to opposites test, 450; to computation, 469 f.; to ink-blot test, 624; to sentence completion, 634; to range of information, 684 f. See General intelligence
Sense-department, effect of, on immediate memory, 555 ff.; on logical memory, 580
Sentence formation, Masselon's method, 627-630; Meumann's method, 630 ff.; Burt's method, 631 f.; Binet's method (sentence completion), 632 ff.
Sex-differences, in description, 378; in report, 398 f.; in uncontrolled association, 413, 415-418; in discrete association test, 430 f.; in part-whole test, 442; in genus-species test, 444; in opposites test, 450; in analogies test, 459; in computation, 469; in mirror-drawing, 491 f.; in substitution, 506; in immediate memory, 545-550; in logical memory, 579; in size-weight illusion, 595; in warmth-illusion, 617; in sentence completion, 634; in word-building, 645; in completion test, 658 f.; in vocabulary, 680; in range of information, 685
Size-weight illusion, 590-598
Social status, relation of, to report, 400
Span tests, see Memory
Specht's test, see Computation

Speed, relation of, to accuracy in computation, 473; in mirror-drawing, 496
Star-test, see Mirror-drawing
Stories, invention of, 634-638
Substitution test, 499-516
Subtraction, see Computation
Suggestibility, tests of, 588-618
Suggestion, by questions, 385 f., 391, 399, 402 f.; of line-lengths, 608-612; by illusion of warmth, 612-618; see Suggestibility
Suggestion-blocks, see Size-weight illusion
Symbol-digit test, see Substitution

Testimony, see Report
Theme, development of, 638 f.; choice of, 639
Transfer of practise, in computation, 471; in mirror-drawing, 491 ff.; in substitution, 509-512; in immediate memory, 551 f., 581; in logical memory, 637.
Types, in descriptions, 380 f.; in work-curves, 476 f.; in logical memory, 585 f.; in ink-blot test, 625 f.; in sentence completion, 633 f.; in invention of stories, 637 f.

Warmth, suggestion of, 612-618
Word-building, 640-649
Work-curve, analysis of, 475-482
Vocabulary test, 640, 674-683

Fatigue 86
115
195